

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 17, 2003, 18:06:54 ; Search time 3970 Seconds
(without alignments)
2947.142 Million cell updates/sec

Title: US-10-054-313-1
Perfect score: 1546
Sequence: 1 MSWLLFLAHRVALAALPCR.....FIGNEADRLAREGAKQSD 286

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-MODEL=frame+ p2n.model -DB=GenEmbl -O=/cpn2_1/uspto spoal/US10054313/runat 17122003 150743 24179/app query.fasta_1.455
-DB=GenEmbl -OFT=fastap -SUFFIX=rge -MINMATCH=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPS=ZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054313 @CGN 1.1 3508 @runat 17122003 150743 24179 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rdt.*
36: em_htg_nam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_hgo_hum.*
40: em_hgo_mus.*
41: em_hgo_other.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result | No. | Score | Match | Length | DB | ID | Description |
|--------|--------|-------|--------|--------|------------|--------------|---------------------|
| 1 | 1526 | 98.7 | 1158 | 9 | BC002973 | Homo sapi | BC002973 Homo sapi |
| 2 | 1526 | 98.7 | 1158 | 9 | AF039652 | Homo sapi | AF039652 Homo sapi |
| 3 | 1526 | 98.7 | 1604 | 9 | AK075490 | Homo sapi | AK075490 Homo sapi |
| 4 | 1522 | 98.4 | 1113 | 9 | AF048994 | Homo sapi | AF048994 Homo sapi |
| 5 | 1522 | 98.4 | 1147 | 9 | AF048995 | Homo sapi | AF048995 Homo sapi |
| 6 | 1518 | 98.2 | 861 | 9 | HSJ4117 | Homo sapi | AJ224117 Homo sapi |
| 7 | 1457.5 | 94.3 | 2120 | 9 | AK096913 | Homo sapi | AK096913 Homo sapi |
| 8 | 1197.5 | 77.5 | 1430 | 10 | AF048993 | Mus muscu | AF048993 Mus muscu |
| 9 | 1193.5 | 77.2 | 1409 | 10 | BC019411 | Mus muscu | BC019411 Mus muscu |
| 10 | 1174 | 75.9 | 178965 | 9 | AC098850 | Homo sapi | AC098850 Homo sapi |
| 11 | 1174 | 75.9 | 192337 | 9 | AC022596 | Homo sapi | AC022596 Homo sapi |
| 12 | 1165 | 75.4 | 211403 | 9 | AC126352 | Homo sapi | AC126352 Homo sapi |
| 13 | 1150.5 | 74.4 | 207418 | 9 | AC090774 | Homo sapi | AC090774 Homo sapi |
| 14 | 1145.5 | 74.1 | 176040 | 9 | AC107926 | Homo sapi | AC107926 Homo sapi |
| 15 | 1016 | 65.7 | 131239 | 9 | HS667H12 | Human DNA | AL035414 Human DNA |
| 16 | 956 | 61.8 | 2024 | 9 | AY063502 | Cercopith | AY063502 Cercopith |
| 17 | 923 | 59.7 | 1992 | 9 | AY063503 | Macaca as | AY063503 Macaca as |
| 18 | 882 | 57.1 | 1293 | 5 | CHKESTFL25 | Gallus gall | D26340 Gallus gall |
| 19 | 816 | 52.8 | 2165 | 9 | AF448495 | Callithri | AF448495 Callithri |
| 20 | 755.5 | 48.9 | 1911 | 9 | AK057473 | Homo sapi | AK057473 Homo sapi |
| 21 | 755.5 | 48.9 | 3471 | 9 | BC041488 | Homo sapi | BC041488 Homo sapi |
| 22 | 423.5 | 27.4 | 1153 | 3 | AY089374 | Drosophil | AY089374 Drosophil |
| 23 | 375 | 24.3 | 1460 | 3 | AF032921 | Drosophil | AF032921 Drosophil |
| 24 | 373.5 | 24.2 | 1117 | 3 | TBU74470 | Trypanosoma | U74470 Trypanosoma |
| 25 | 373.5 | 24.2 | 141775 | 2 | AC105378 | Trypanoso | AC105378 Trypanoso |
| 26 | 356.5 | 23.1 | 37983 | 2 | AC020394 | Drosophil | AC020394 Drosophil |
| 27 | 356.5 | 23.1 | 85651 | 2 | AC005463 | Drosophil | AC005463 Drosophil |
| 28 | 356.5 | 23.1 | 86889 | 3 | AC005448 | Drosophil | AC005448 Drosophil |
| 29 | 356.5 | 23.1 | 139410 | 3 | AC007084 | Drosophil | AC007084 Drosophil |
| 30 | 356.5 | 23.1 | 234023 | 3 | AE003839 | Drosophil | AE003839 Drosophil |
| 31 | 331.5 | 21.4 | 34828 | 3 | U41994 | Caenorhabd | U41994 Caenorhabd |
| 32 | 312 | 20.2 | 166143 | 2 | AL356960 | Homo sapi | AL356960 Homo sapi |
| 33 | 303.5 | 19.6 | 2277 | 3 | TRFRNH1A | Crithidia f | LI8916 Crithidia f |
| 34 | 300.5 | 19.4 | 56083 | 9 | AC108488 | Homo sapi | AC108488 Homo sapi |
| 35 | 276 | 17.9 | 39512 | 8 | SPBC336 | S. pombe | AL121815 S. pombe |
| 36 | 272 | 17.6 | 1557 | 8 | AF048992 | Schizosac | AF048992 Schizosac |
| 37 | 272 | 17.6 | 119211 | 6 | AX067465 | Sequence | AX067465 Sequence |
| 38 | 268 | 17.3 | 231 | 6 | BD118436 | EST and e | BD118436 EST and e |
| 39 | 261.5 | 16.9 | 207104 | 2 | AC136986 | Mus muscu | AC136986 Mus muscu |
| 40 | 258 | 16.7 | 107893 | 5 | BX276180 | Zebratfish | BX276180 Zebratfish |
| 41 | 255.5 | 16.5 | 231062 | 2 | AC109110 | Rattus no | AC109110 Rattus no |
| 42 | 255.5 | 16.5 | 246539 | 2 | AC125638 | Rattus no | AC125638 Rattus no |
| 43 | 254 | 16.4 | 683 | 3 | AF542055 | Leishmani | AF542055 Leishmani |
| 44 | 254 | 16.4 | 40397 | 8 | SC9599 | S. cerevisia | Z49999 S. cerevisia |
| 45 | 253 | 16.4 | 300600 | 1 | AF005369 | Thermosyn | AF005369 Thermosyn |

ALIGNMENTS

RESULT 1

```

BC002973      1156 bp      mRNA      linear      PRI 12-JUL-2001
LOCUS      Homo sapiens, ribonuclease H1, clone MGC:2019 IMAGE:3537074, mRNA,
DEFINITION      complete cds.
ACCESSION      BC002973
VERSION      BC002973.1 GI:12804228
KEYWORDS      MGC.
SOURCE      Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 1156)
AUTHORS      Strausberg, R.
TITLE      Direct Submission
JOURNAL      Submitted (06-FEB-2001) National Institutes of Health, Mammalian
      Gene Collection (MGC), Cancer Genomics Office, National Cancer
      Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
      USA
REMARK      NIH-MGC Project URL: http://mgi.nci.nih.gov
COMMENT      Contact: MGC help desk
      Email: cgapbs@mail.nih.gov
      Tissue Procurement: DCTP/DTP
      CDNA Library Preparation: Rubin Laboratory
      CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
      DNA Sequencing by: Genome Sequence Centre,
      BC Cancer Agency, Vancouver, BC, Canada
      info@bcgsc.bc.ca
      Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
      Susanna Chan, Readman Chiu, Chris Fiehl, Erin Garland, Ran Guin,
      Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
      Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
      Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
      Schein, Duane Smalish, Michael Smith, Lorraine Spence, Jeff Stott,
      Michael Thorne, Miranada Teai, Natasja van den Bosch, Jill Vardy,
      George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 8 Row: 1 Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 3004980.

FEATURES             Location/Qualifiers
     source            1..1158
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="LocusID:6040"
                     /db_xref="taxon:9606"
                     /clone="MGC:2019 IMAGE:3537074"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_1ib="NIH MGC 7"
                     /lab_host="DH10B-R"
                     /notes="Vector: pOT7"
                     71..931
                     /codon_start=1
                     /product="ribonuclease H1"
                     /protein_id="AAH02973.1"
                     /db_xref="GI:12804229"
                     /translation="MSKLLFLAHVALALPCRGSRGPFMEFVVRGKTKGTWLN
      ECAQVDRFPARFKFAIEDENAFVRKSAPEVSEGHENGHGSEKAKRLREP
      LDGDGHESAPFAGKMKPSPVAPPVSRDTFTSMDGDFVVTYDGCSSNGRRPRAGI
      GYVNGFCHPLNGVIRLPGTNQRAIEHAACKAIEQAKTINKLVLYTDSMTFNGI
      TNVVGKNGKNGTWSAGKEVINKEDFVALERLTQGMIDIQMHPVGHSGFIGNEADRL
      AREGAKQSED"
BASE COUNT      317 a      248 c      349 g      244 t
ORIGIN
Alignment Scores:
Pred. No.:      1.61e-101      Length:      1156
Score:      1526.00      Matches:      282
Percent Similarity:      99.30%      Conservative:      2
Best Local Similarity:      98.60%      Mismatches:      2
Query Match:      98.71%      Indels:      0
DB:      Gaps:      0

```

```

US-10-054-313-1 (1-286) x BC002973 (1-1158)
Qy      1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db      71 ATGAGCTGGCTTCTGTTCCTGGCCACAGAGTCGCTTGGCCGCTTGGCCGCGC 130
Qy      21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgGlyThrGlyValPhe 40
Db      131 GGCTCTCGGGGTTTCGGGATGTTCTATGCGGTGAGGAGGGCGCGCAAGACCGGGTCTTT 190
Qy      41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db      191 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCTCTGCTGCCAGATTTAAGAG 250
Qy      61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db      251 TTTGCCACAGAGATGAGGCTGGGCTTTTGTGAGGAATCTGCAAGCCGCGGAGTTTCA 310
Qy      81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db      311 GAAGGGCATGAAATCAACATGGACAGAATCGAGGCGGAAAGCCAGCAGACTCCGT 370
Qy      101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db      371 GAGCCACTGGATGGAGTGGACATGAAGCGGAGCGCGGTATGCCAAGGACATGAAGCCG 430
Qy      121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db      431 AGCGTGGAGCGCGCGCTCCAGTTAGCAGACACGCTTTCTTACATGGAGAGACTTCGTC 490
Qy      141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160
Db      491 GTCGCTACACTGATGCTGCTCTCCAGTAAATGGGCGTAGAAGCGCGGAGCAGGAATC 550
Qy      161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLysArgLeuProGlyArgGln 180
Db      551 GCGCTTACTGGGGCGCAGGCCATCCTTAAATGTAGGCATTAGACTTCTCTGGCGGCGAG 610
Qy      181 ThrAsnGlnArgAlaGluLeuHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
Db      611 ACAACCAAGAGCGGAAATTCATGCGCTGCAAGCCATTGAACAGCAAGCAAGACTCAA 670
Qy      201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyLysThrAsn 220
Db      671 AACATCAATAAATGCTGTTCTGTATACAGACAGTATGTTTACGATAAATGCTATAACTAC 730
Qy      221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValLysAsn 240
Db      731 TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGCAAGTGCAGGGAAGAGGTGATCAAC 790
Qy      241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db      791 AAGAGGACTTTGTGGCACTGGAGAGGCTTACCAGGGGATGGACATTCAGTGGATGCAT 850
Qy      261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db      851 GTTCTGCTCATTCGGGATTTATAGGCAATCAAGAGTGCACAGATTAGCAGAGAAGGA 910
Qy      281 AlalysGlnSerGluAsp 286
Db      911 GCTAAACAATCGGAAGAC 928

```

```

RESULT 2
AF039652      1168 bp      mRNA      linear      PRI 02-APR-1998
LOCUS      Homo sapiens ribonuclease H type II mRNA, complete cds.
DEFINITION      AF039652
ACCESSION      AF039652.1 GI:3004980
VERSION      AF039652.1
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1168)
Wu, H., Lima, W. and Crooke, S.
Molecular cloning and expression of cDNA for human RNase H
Antisense Nucleic Acid Drug Dev. (1998) In press
2 (bases 1 to 1168)
Wu, H., Lima, W. and Crooke, S.
Direct Submission
Submitted (22-DEC-1997) Molecular Pharmacology, Isis
Pharmaceuticals, Inc., 2292 Faraday Ave, Carlsbad, CA 92008, USA
Location/Qualifiers
1. 1168
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p11.2"
82. 942
/function="ribonuclease specific for the RNA strand of a
DNA-RNA hybrid"
/note="RNase HII"
/codon_start=1
/product="ribonuclease H type II"
/protein_id="AAC09261.1"
/db_xref="GI:3004981"
/translation="MSWLLFLAHRVALAALPCRRSGRFGMFYAVRRGRKTVFLTW
ECRAVDPPAPARPKFATDEAWAFVKKSASPEVSEGHQNGHGESEAKSKRLREP
LDGDGESAEPVAKHMKSVBPAPVSDTFVMDGVVYVTDCCSNGRRPRAGI
GYWGPGLNVLGRIPROTNQRAIHAACKAEQAKTONKLVLYTDSMPFINGI
TNWQGVKNGKWSAGKEVINKEDFVALERTQGMIDIQMHWVPHSGFIGNERDL
AREGAKQSD"

BASE COUNT 325 a 249 c 349 g 245 t
ORIGIN

Alignment Scores:
Pred. No.: 1,628-101 Length: 1168
Score: 1526.00 Matches: 282
Percent Similarity: 99.30% Conservative: 2
Best Local Similarity: 98.60% Mismatches: 2
Query Match: 98.71% Indels: 0
DB: 9 Gaps: 0

US-10-054-313-1 (1-286) x AF039652 (1-1168)

Qy 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArg 20
Db 82 ATGAGCTGCTTCTGTCTCGCCACACAGATCGCTTGGCGCGCTTGCCTCGCGCGC 141
Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgGlyThrGlyValPhe 40
Db 142 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGAGGGCGCCGACACCGGGTCTTT 201
Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60
Db 202 CTGACCTGAATGAGTGCAGACACAGTGGACCGGTTCTCTGCTGCCAGATTTAAGAAG 261
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 262 TTGGCCACAGAGATGAGCGCTTGGCGCTTTGTCAGAAATCTGCAAGCCGGAAGTTCA 321
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 322 GAAGGCGATGAAATCAACATGCAGCAAGATCGGAGGCGAAGCCAGCAGCGACTCGGT 381
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 382 GAGCCACTCGATGGAGATGACATGAAGCGCAGAGCCGCTATGCAAAACACATGAAGCGC 441
Qy 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 442 ACCGTGGAGCCGGCGCTCCAGTTAGCAGACACAGTTTCTCTACATGGGAGCTTCGTC 501
Qy 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160

Db 502 GTGCTTACACTGATGGCTGCTCCAGTAATGGCGTAGAAGCGCGCAGCAGAAATC 561
Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db 562 GCGGTTTACTGGGGGCCAGGCCATCTTTAAATGTAGGCATTAGACTTCTCTGGCGGCAG 621
Qy 181 ThrAsnGluArgNlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
Db 622 ACAAAACAAAGAGCGAAATTCATGCAGCTCGCAAGCCATTGAAACCAAGCAAGACTCAA 681
Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 682 AACATCAATAAAGCTGTTCTGTATACAGACAGATGTTTACGATAAATGGTAACTAAC 741
Qy 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
Db 742 TGGGTTCAAGGTTTGAAGAAAATGGGTGGAACAGCAAGTGCAGGGAAGAGGTGTATCAAC 801
Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db 802 AAAGAGGACTTTGTGGCACTGGAGAGCTTACCAGGGGATGGACATTCACTGGATGCAT 861
Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db 862 GTTCTCGTCAATTCGGGATTTATAGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA 921
Qy 281 AlaLysGlnSerGluAsp 286
Db 922 GCTAAACATCGGAAGAC 939

RESULT 3
AK075490 1604 bp mRNA linear PRI 03-SEP-2002
LOCUS Homo sapiens cDNA PSEC0185 fis, clone OVARC1002091, highly similar
to Homo sapiens ribonuclease H type II mRNA.
AK075490
ACCSSION AK075490.1 GI:22761667
VERSION AK075490.1 GI:22761667
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

Ota, T., Nishikawa, T., Suzuki, Y., Kawai-Hio, Y., Hayashi, K.,
Ishii, S., Saito, K., Yamamoto, J., Wakamatsu, A., Nagai, T.,
Nakamura, Y., Nagahara, K., Sugano, S. and Isogai, T.
HRI human cDNA sequencing project
Unpublished
2 (bases 1 to 1604)
Isogai, T. and Yamamoto, J.
Direct Submission
Submitted (20-MAR-2002) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
HRI human cDNA sequencing project; cDNA 5'- & 3'-end one pass
sequencing, clone selection and full insert sequencing; Helix
Research Institute (supported by Japan Key Technology Center etc.);
cDNA library construction: Institute of Medical Science, University
of Tokyo, Laboratory of Genome Structure, Human Genome Center.
1. 1604
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="OVARC1002091"
/tissue_type="ovary, tumor tissue"
/clone_lib="OVARC1"
/notes="cloning vector: pME189PL3"
BASE COUNT 426 a 343 c 460 g 375 t
ORIGIN

Alignment Scores:

Pred. No.: 2,31e-101 Length: 1604
 Score: 1526.00 Matches: 282
 Percent Similarity: 99.30% Conservative: 2
 Best Local Similarity: 98.60% Mismatches: 2
 Query Match: 98.71% Indels: 0
 DB: 9 Gaps: 0

US-10-054-313-1 (1-286) x AK075490 (1-1604)

Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
 Db 61 ATGAGCTGGCTTCTGCTGCGCCACAGAGTCGCTTGGCGCGCTTGGCTGCGCGCGC 120
 Qy 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgArgGlyValGlyThrGlyValPhe 40
 Db 121 GCTCTCCGGGTTCCGATGTTCTATCCCTGAGAGGCGCCGACACCGGGTCTTT 180
 Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
 Db 181 CTGACCTGGATGATGTCAGACAGACAGTGCACCGCTTCTGCTGCGAGATTTAAGAAG 240
 Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
 Db 241 TTTCACACAGAGATGAGCGCTTGGCGCTTGTTCAGGAAATCTGACAGCCCGAGATTCA 300
 Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLysArg 100
 Db 301 GAAGGCGATGAATCAACATGCACAGATCGGAGCGGAAAGCAGCAGCGACTCCGT 360
 Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTrpAlaLysHisMetLysPro 120
 Db 361 GAGCACCTGGATGGAGATGGACATGAAGCGGAGCGCGTATGCAAGACATGAAGCGG 420
 Qy 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
 Db 421 ACGTGGAGCGCGCTCCAGTTAGCAGACAGACAGTCTTCTCATCGGAGACTTCGTC 480
 Qy 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyTle 160
 Db 481 GTCCGTACACTGATGGCTGCTGCTCCAGTAAATGGCGGTAGAGCGCGGAGGAATC 540
 Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
 Db 541 GCGCTTACTGGGGCGCGGCGCATCTCTTAAATGATAGGCACTAGACTTCTCGCGCGGAG 600
 Qy 181 ThrAsnGlnArgAlaGluIleHisAlaLysLysAlaIleGluGluAlaLysThrGln 200
 Db 601 ACAAAACCAAGACGCGAAATTCATGCGCTGCAAGGCCATTTGAACAACAAAGACTCAA 660
 Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyTleThrAsn 220
 Db 661 AACATCAATAACTGGTCTGTATACAGACAGATATGTTACAGATTAATGGTATACTAAC 720
 Qy 221 TrpValGlnGlyTrpLysLysAsnGlyTyrLysThrSerAlaGlyLysGluValIleAsn 240
 Db 721 TGGGTTCAAGTTGGAAGAAATAATGGTGAAGACAGTGCAGGGAAGAGGTGATCAAC 780
 Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
 Db 781 AAGAGGACTTTGTGGCATGGAGAGCTTACCAGGGGATGGACATTCAGTGGATGAT 840
 Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
 Db 841 GTTCCTGGTCAATTCGGGATTTATAGCAATGAAGAGCTGACAGATTAGCCAGAGAAGA 900
 Qy 281 AlaLysGlnSerGluAsp 286
 Db 901 GCTAAACCAATCGGAGAC 918

RESULT 4
 AF048994
 LOCUS Homo sapiens ATCC158373 ribonuclease H1 (rnhl) mRNA, complete cds.
 DEFINITION

AF048994
 AF048994.1 GI:2935441
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Cerritelli,S.M., Fedoroff,O.Y., Reid,B.R. and Crouch,R.J.
 A common forty amino acid motif in eukaryotic RNases H1 and
 caulimovirus ORF VI proteins binds to duplex RNAs
 Nucleic Acids Res. (1998) in press
 2 (bases 1 to 1113)
 Cerritelli,S.M. and Crouch,R.J.
 Cloning, expression, and mapping of ribonucleases H of human and
 mouse related to bacterial RNase H1
 Genomics 53 (3), 300-307 (1998)
 99017966
 979596
 3 (bases 1 to 1113)
 Cerritelli,S.M. and Crouch,R.J.
 Direct Submission
 Submitted (17-FEB-1998) Laboratory of Molecular Genetics, NICHD
 NIH, 6 Center Drive Building 6B Room 2B-231, Bethesda, MD 20892,
 USA

Location/Qualifiers

1. .1113
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /strain="ATCC158373"
 /db_xref="ATCC:158373"
 /db_xref="dbEST:AA356084"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17p11.2"
 /cell_type="Jurkat T-cells VI"
 1. .1113
 /gene="rnhl"
 28. .888
 /genes="rnhl"
 /notes="similar to bacterial RNases H1"
 /codon_start=1
 /product="ribonuclease H1"
 /protein_id="AAC78563.1"
 /db_xref="GI:2935442"
 /translation="MSWFLAHARVALAALPCRSGRGFGMFVAVRRGRTGVFLTWN
 ECRQVDRPDAARFKFAFEDENAWFVKASPEVSECHENGCHGDESEAKKRLSEP
 LDGDGHESAPPYAKHKMPSVEFAPVSRDITFIMGDFVYVYDCCSSNGRRPRAGI
 GYVWGFHPLNVGIRLPGRQTNORAEIHAACKAIEQAKTNINKLVITDTSFTINGI
 TNWVGKNGKNGKTSAGKEVINKEFVALERLTQGMDIQNMHVPHSGHGFIGNEEADRL
 AREGAKQSD"

gene

CDS

BASE COUNT 311 a 232 c 333 g 237 t

ORIGIN

Alignment Scores:

Pred. No.: 2,99e-101 Length: 1113
 Score: 1522.00 Matches: 281
 Percent Similarity: 98.95% Conservative: 2
 Best Local Similarity: 98.25% Mismatches: 3
 Query Match: 98.45% Indels: 0
 DB: 9 Gaps: 0

US-10-054-313-1 (1-286) x AF048994 (1-1113)

Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
 Db 28 ATGAGCTGGTTCGTTCCTTGGCCACAGATCGCTTGGCGCTTGGCTGCGCGC 87
 Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyLysThrGlyValPhe 40
 Db 88 GCGCTCGCGGGTTCGGGATGTTCTATCGCGTGAGGAGGCGCGCAAGACCGGGGCTTT 147
 Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60


```

Db 148 CTGACCTGGAATGATGAGCAGCAGAGTGGACCGTTCTCTGCTGCCAGATTTAAGAAG 207
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 208 TTTGCCACAGAGGATGAGCGCTGGCCCTTTGTTCAGGAATCTGCAAGCGCGGAAGTTTCA 267
Qy 81 GluGlyHisGluAsnGlnHisGlyClnGlnLysSerGluAlaLysProGlyLysArgLeuArg 100
Db 268 GAAGGGCATGAAATCAACATGGACAGAATCGGAGGCGAAGCCAGCAGACTCCGT 327
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 328 GAGCCACTGGATGGAGATGGACATGAAGCGCAGAGCGGTATGCAAGCACATGAGCCG 387
Qy 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 388 AGCGTGGAGCGCGCCCTCCAGATTAGCAGACACAGTTTCTTCTACATGGAGACTTCGTC 447
Qy 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgLysProArgAlaGlyIle 160
Db 448 GTCGCTACATGATGGCTGCTCCAGTATGGCGGTAGAGCGCCGCGAGGATC 507
Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db 508 GCGCTTTACTGGGGCCAGGCCCATCTTTAAATGTAGGCATTAGACTTCTCTGGCGCGAG 567
Qy 181 ThrAsnGlnArgAlaGluIleHisAlaCysLysAlaIleGluGlnAlaLysThrCln 200
Db 568 ACAAAACCAAGAGCGGAATTCATGACCTGCAAGCCATTGAACAAGCAAGACTCAA 627
Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 628 AACATCAATAAATGCTGCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC 687
Qy 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
Db 688 TGGGTTCAAGTGTGAAGAAATCGGTGGAAGCAAGTGCAGGGAAGAGGTGATCAAC 747
Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db 748 AAAGAGGACTTTGTGGCCTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 807
Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db 808 GTTCTGTGTCATTCGGGATTTATAGCAATGAAGAAGCTGACAGATTAGCCAGAGAGGA 867
Qy 281 AlaLysGlnSerGluAsp 286
Db 868 GCTAAACAATCGGAAGAC 885

RESULT 5
AF048995 1147 bp mRNA linear PRI 21-NOV-1998
LOCUS Homo sapiens ATCC159806 ribonuclease H1 (rnhl) mRNA, complete cds.
DEFINITION AF048995
ACCESSION AF048995
VERSION AF048995.1 GI:2935443
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1147)
Cerritelli,S.M., Fedoroff,O.Y., Reid,B.R. and Crouch,R.J.
A common forty amino acid motif in eukaryotic RNases H1 and
caulimovirus ORF VI proteins binds to duplex RNAs
Nucleic Acids Res. (1998) In press
2 (bases 1 to 1147)
Cerritelli,S.M. and Crouch,R.J.
Cloning, expression, and mapping of ribonucleases H of human and
mouse related to bacterial RNase H1
Genomics 53 (3), 300-307 (1998)
99017966

```

```

9799596
3 (bases 1 to 1147)
Cerritelli,S.M. and Crouch,R.J.
Direct Submission
Submitted (17-FEB-1998) Laboratory of Molecular Genetics, NICHD
NIH, 6 Center Drive Building 6B Room 2B-231, Bethesda, MD 20892,
USA
FEATURES
Location/Qualifiers
1. .1147
/organism="Homo sapiens"
/mol_type="mRNA"
/strain="ATCC159806"
/db_xref="ATCC:159806"
/db_xref="dbEST:AA312554"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p11.2"
/cell_type="Jurkat T-cells VI"
1. .1147
/gene="rnhl"
2. .922
/gene="rnhl"
/notes="similar to bacterial RNases H1"
/codon_start=1
/product="ribonuclease H1"
/protein_id="AAC78564.1"
/db_xref="GI:2935444"
/tranlation="MSWFLFLAHRVALAALPCRGRSGRGMFVAVRRGRKTGVFLTN
EERAVDRPFAARFKFKATEDENAFVRKASPEYSEGHENQHGQSEAKSKARLEP
LQDGESAEPAKMKPSVEFAPPVSRDTFSIGDFVVTYIDGCCSNRRPRAGI
GVYWGKHLNIGIRLPGRNORAHAAKCAIKSOAKTONKINLVLYTDSMTPTNGI
TNWQWQKNGKWTSGKEVINKEDFVALERLTQGMNDIQMHEVPGHSGFIGNEADRL
AREAKQSD"
BASE COUNT 316 a 244 c 345 g 242 t
ORIGIN
Alignment Scores:
Pred. No.: 3,09e-101 Length: 1147
Score: 1522.00 Matches: 281
Percent Similarity: 98.95% Conservative: 2
Best Local Similarity: 98.25% Mismatches: 3
Query Match: 98.45% Indels: 0
DB: 9 Gaps: 0
US-10-054-313-1 (1-286) x AF048995 (1-1147)
Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 62 ATGAGCTGGTTTCTGTTCTCTGGCCACAGAGTGCCTTGGCCGCTTGCCTGCCGCCG 121
Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgLysThrGlyValPhe 40
Db 122 GGCTCTCGCGGTTCGGGATGTTCTATGCCGTGAGGAGGGCGCGAAGCCGGGTCTTT 181
Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db 182 CTGACCTGGAAATGAGTGCAGAGCAGAGTGGACCGGTTTCTCTGCTGCCAGATTAGAAG 241
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 242 TTTGCCACAGAGGATGAGGCTGGGCTTGTTCAGGAATCTCAGAACCCCGGAAGTTTCA 301
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGlnLysSerGluAlaLysProGlyLysArgLeuArg 100
Db 302 GAAGGGCATGAAATCAACATGGACAGAATCGGAGGCGAAGCCAGCAGACTCCGT 361
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 362 GAGCCACTGGATGGAGATGGACATGAAGCGCAGAGCGGTATGCAAGCACATGAGCCG 421
Qy 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 422 AGCGTGGAGCGCGCCCTCCAGTTAGCAGACACCGTTTCTTCTACATGGAGACTTCGTC 481

```

QY 141 ValValThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
 DB 482 GTGGCTACACTGATGGCTGTCTCCAGTAATGGCGCTAGAGCGCGCAGCAGGAATC 541
 QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
 DB 542 GCGGTTTACTGGGGCCAGGCATCTTTAAATGTAGSCATAGACTTCTTGGGGCGCAG 601
 QY 181 ThrAsnGlnArgAlaGluHisAlaCysLysAlaIleGluGlnAlaLysThrGln 200
 DB 602 ACAACCAAGAGCGGAATTCATCGACCTGCAAGGCAATGAACAAGCAAGACTCAA 661
 QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
 DB 662 AACATCAATAACTGGTCTGTATACAGACAGTATGTTTACGATAATGGTATCACTAAC 721
 QY 221 TrpValGlnGlyTrpLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
 DB 722 TGGGTTCAAGGTTGGAAGAAAATCGGTGAAGACAAAGTGCAGGMAAGAGGTGATCAAC 781
 QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
 DB 782 AAAGAGGACTTTGTGGCTGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGAT 841
 QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
 DB 842 GTTCTGTGTCATTCGGATTATAGCAATGAAGAAGCTGCACAGATTAGCCAGAGAGGA 901
 QY 281 AlaLysGlnSerGluAsp 286
 DB 902 GCTAAACCAATCGGAAGAC 919
 RESULT 6
 HSAJ4117 861 bp mRNA linear PRI 07-APR-1998
 LOCUS HSAJ4117
 DEFINITION Homo sapiens mRNA RNase HII.
 ACCESSION AJ224117
 VERSION AJ224117.1 GI:3043446
 KEYWORDS RNase HII; rnhii Gene.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 861)
 AUTHORS Frank, P., Braunschöfer-Reiter, C. and Wintersberger, U.
 TITLE Cloning and functional expression of human RNase HII
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 861)
 AUTHORS Frank, P., Braunschöfer-Reiter, C. and Wintersberger, U.
 TITLE Direct Submission
 JOURNAL Submitted (22-FEB-1998) Department of Molecular Genetics, Institute
 of Tumor Biology-Cancer Research, Borschkegasse 8a, Vienna, Austria
 A-1090
 FEATURES
 Location/Qualifiers
 1..861
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 1..861
 /gene="rnhii"
 1..861
 /gene="rnhii"
 /note="evolutionarily related to prokaryotic RNase HI and
 RNase H domain of retroviral reverse transcriptase;
 ribonuclease Hii"
 /codon_start=1
 /product="RNase HII"
 /protein_id="CA11835.1"
 /db_xref="GI:3043447"
 /db_xref="SPTREMBL:O60857"
 /translation="MSWLLFLAHRVALAALPCRGRSGRFMEYAVRRKGTGVFLTWN
 ECRQVDRFPAPAFKKFATDEAWAFVRKSASPEVSEGHENQHGRESAKAKRLREP

LDGCHSENEPVAKHKPSVBPAPVSRDTSYMGDFVVVYVTDGCCSNRRRPPAGI
 GYWGFGPLNVLPGQRQNRRAEHAACKAEQAKTONIKLVLYTDSNFTINGI
 TWWVGKKNKGNKTSAGREVINKDFVALERLTQGMNDIQMHVPHGHSFIGNEADRL
 AREGAKQSD"
 BASE COUNT 236 a 189 c 263 g 173 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4,37e-101 Length: 861
 Score: 1518.00 Matches: 280
 Percent Similarity: 99.30% Conservative: 4
 Best Local Similarity: 97.90% Mismatches: 2
 Query Match: 98.19% Indels: 0
 DB: 9 Gaps: 0
 US-10-054-313-1 (1-286) x HSAJ4117 (1-861)
 QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
 DB 1 ATGAGCTGGCTTCTGTCTCTGCCCCACAGAGTCGCTTGGCCGCCCTTGGCCCTGCCGCCGC 60
 QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgAtcGlyArgLysThrGlyValPhe 40
 DB 61 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTAGAGGGGGCGCAAGACCGGGGTCTTT 120
 QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
 DB 121 CTGACCTGGAATGAGTGCAGAGCACAAAGTGGACCGGTTTCTGCTGCCAGATTTAAGAAG 180
 QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
 DB 181 TTTCACACAGAGATGAGGCTGGGCTTTGTGAGGAAATCTGCAAGCCGCGGAAGTTTCA 240
 QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
 DB 241 GAAGGCGATGAATCAACATGACGAGAGATCGAGCGCAAGCCAGCAGACTCCGT 300
 QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
 DB 301 GAGCCACTGGATGGATGGACATGAAGCGCAGAGCCGTATGCAAGACACATGAAGCGCG 360
 QY 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
 DB 361 AGGTGGAGCGCGGCTCCAGTTACAGACACACGTTTCTACATGGAGACTTCGTC 420
 QY 141 ValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
 DB 421 GTGCTTACACTGATGGCTGCTCCAGTAATGGCGCTAGAGCGCGCAGCAGGAATC 480
 QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
 DB 481 GCGGTTTACTGGGGCGCAGGCCATCTTTAAATGTAGGCAATAGACTTCTTGGCGCGCAG 540
 QY 181 ThrAsnGlnArgAlaGluIleHisAlaCysLysAlaIleGluGlnAlaLysThrGln 200
 DB 541 ACAAAACCAAGAGCGGAATTCATCGACCTGCAAGGCAATGAACAAGCAAGACTCAA 600
 QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
 DB 601 AACATCAATAACTGGTCTGTATACAGACAGTATGTTTACGATAATGGTATCACTAAC 660
 QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
 DB 661 TGGGTTGAGGTTGGAAGAAAATCGGTGAAGACAAAGTGCAGGMAAGAGGTGATCAAC 720
 QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
 DB 721 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGAT 780
 QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
 DB 781 GTTCTGTGTCATTCGGATTATAGCAATGAAGAAGCTGCACAGATTAGCCAGAGAGGA 840

```

Qy 281 AlalysGlnSerGluasp 286
|||||
Db 841 CTAACCAATCGGAGAC 856

RESULT 7
AK096913 2120 bp mRNA linear PRI 15-JUL-2002
LOCUS Homo sapiens cDNA FLJ39594 fis, clone SKNSH2001875, highly similar
DEFINITION to Homo sapiens ATCC158373 ribonuclease H1 (rhnl) mRNA.
ACCESSION AK096913
VERSION AK096913.1 GI:21756512
KEYWORDS cllgo capping, fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Kawakami, B., Sugiyama, A., Takemoto, M., Sugiyama, T., Irie, R.,
Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,
Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,
Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,
Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A.,
Oshima, A., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K.
and Isogai, T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2120)
AUTHORS Isogai, T. and Yamamoto, J.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kanetari, Kisarazu, Chiba 292-0812, Japan, Fax: 81-438-52-3986)
(E-mail: genomics@nri.co.jp, Tel: 81-438-52-3975,
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5' & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
source
1..2120
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="SKNSH2001875"
/cell_line="SK-N-SH"
/clone_lib="SKNSH2"
/note="cloning vector: pME18SFL3"
BASE COUNT 546 a 476 c 624 g 474 t
ORIGIN

Alignment Scores:
Pred. No.: 2,77e-96 Length: 2120
Score: 1457.50 Matches: 282
Percent Similarity: 77.60% Conservative: 2
Best Local Similarity: 77.05% Mismatches: 2
Query Match: 94.28% Indels: 80
DB: 9 Gaps: 1

US-10-054-313-1 (1-286) x AK096913 (1-2120)

Qy 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 30 ATGAGCTGGCTTCTGTTCTGGCCACAGAGTGGCTTGGCCGCTTGGCTGGCCGCGC 89

Qy 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgArgGlyArgGlyThrGlyValPhe 40
Db 90 GGCTCTGGGGTTCGGGATGTTCTATGCCGTGAGAGGGGGCCGACAGCCGGGTCTTT 149

Qy 41 LeuThrTrp----- 43
|||||

```

RESULT 8
 AF048993
 LOCUS
 DEFINITION Mus musculus ribonuclease H1 (Rnh1) mRNA, complete cds.
 ACCESSION AF048993
 VERSION AF048993.1 GI:2935417
 KEYWORDS

```

Db 150 CTGACCTGCTGAGTGGCCCGCCCGGAGGCGCCTCCTCGCGGCATCGGGTTCGAGGA 209
Qy 43 ----- 43
Db 210 GGGCGGAGCCGGCTAAACCCCGGCGGAGCCTTGGCGGTCTGTGCGCGCGCG 269
Qy 43 ----- 43
Db 270 TGTGCTGTGGTGGTGGGCGGAGCCCTCGTTTCTCTGCTGGCGGGGGTTCAGG 329
Qy 43 ----- 43
Db 330 AGGCTCATGGCTGGGACTGGGAACGCCAAGACCTGTTTGAATATCAGAAACCTCGC 389
Qy 44 -----AsnGluCysArgAlaGlnValAspArgPheProAlaAlaAArgPheLysPh 61
Db 390 GTAAGGAATGAGTGCAGAGCACAGGTGGCCGGTTTCTGCTGCCAGATTTAAGAAGTT 449
Qy 61 eAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSerGI 81
Db 450 TGGCACAGAGGATGAGGCTGGGCTTTGTCAGGAAATCTGCAAGCCCGAAGTTTCAG 509
Qy 81 uGlyVHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGI 101
Db 510 AGGCGATGAATCAACATGGACAGATCGGAGGCGAAAGCCAGCAAGCGACTCCGTGA 569
Qy 101 uProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysProSe 121
Db 570 GCCACTGGATGGAGATGGACATGAAAGCGCAGAGCGGTATGCAAGACACATGAAGCGG 629
Qy 121 rValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheValVa 141
Db 630 CGTGGAGCGCGGCCCTCCAGTTAGCAGACACAGTTTCTCATCGGAGACTTCGTCTGT 689
Qy 141 lValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyLeGI 161
Db 690 CGTCTACACTGATGGCTGCTGCCAGTAATGGCGCTAGAAAGCCCGCAGCAGAAATCG 749
Qy 161 yValTyrTrpGlyProGlyHisProLeuAsnValGlyLleArgLeuProGlyArgGlnTh 181
Db 750 CGTTTACTGGGGCGCGGCCATCTTTAAATGTAGGCATTAGACTTCTCGGGCGGAGAC 809
Qy 181 rAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAs 201
Db 810 AAACCAAGAGCGGAAATTCATGCAGCCTGCAAGCCATTGAACAGCAAGAAAGACTCAAA 869
Qy 201 nIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyLleThrAsnTr 221
Db 870 CATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAATCG 929
Qy 221 pValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLY 241
Db 930 GGTTCAGGTTGGAAGAAAATGGGTGGAAGACAGTGCAGGGAAGAGGTGATCAACAA 989
Qy 241 sGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisVa 261
Db 990 AGAGGACTTTGTGGCCTGGAGAGGCTTACCCAGGGGATGGACATTCACTGGATGTCATGT 1049
Qy 261 lProGlyHisSerGlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGluGlyAl 281
Db 1050 TCCTGGTCATTGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGAGGC 1109
Qy 281 aLysGlnSerGluAsp 286
Db 1110 TAAACATCGGAAGAC 1125

```

SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1430)
 AUTHORS Cerritelli, S.M., Pedoroff, O.Y., Reid, B.R. and Crouch, R.J.
 TITLE A common forty amino acid motif in eukaryotic RNases H1 and caulimovirus ORF VI proteins binds to duplex RNAs
 JOURNAL Nucleic Acids Res. (1998) In press
 REFERENCE 2 (bases 1 to 1430)
 AUTHORS Cerritelli, S.M. and Crouch, R.J.
 TITLE Cloning, expression, and mapping of ribonucleases H of human and mouse related to bacterial RNase H1
 JOURNAL Genomics 53 (3), 300-307 (1998)
 MEDLINE 99017966
 PUBMED 9799596
 REFERENCE 3 (bases 1 to 1430)
 AUTHORS Crouch, R.J. and Cerritelli, S.M.
 TITLE RNases H of lower eukaryotes: Saccharomyces cerevisiae, Schizosaccharomyces pombe, Neurospora crassa and Crithidia fasciculata
 JOURNAL (in) Crouch, R.J. and Toulme, J.-J. (Eds.); RIBONUCLEASES H; INSERM Editions, Paris, France (1998) In press
 REFERENCE 4 (bases 1 to 1430)
 AUTHORS Cerritelli, S.M. and Crouch, R.J.
 TITLE Direct Submission
 JOURNAL Submitted (12-FEB-1998) Laboratory of Molecular Genetics, NICHD NIH, 6 Center Drive, Building 6B, Room 2B-231, Bethesda, MD 20892, USA

FEATURES
 source
 1..1430
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="dbEST:AA163387"
 /db_xref="taxon:10090"
 /chromosome="12"
 /map="12A3"
 /tissue_type="carcinoma"
 /clone_lib="Stratagene 937317"
 /dev_stage="embryonic"
 1..1430
 /gene="Rnh1"
 1..49
 /gene="Rnh1"
 49..906
 /gene="Rnh1"
 /notes="similar to bacterial RNases H1"
 /codon_start=1
 /product="ribonuclease H1"
 /protein_id="AAC78562.1"
 /db_xref="GI:2935418"
 /translations="MRWLPLSLTAVLVRLRRGICGLMFYAVRRGRTGTGFLWS
 ECKAQVDFRPAARFKFPAVEDAWFVRSSSPDGSGQSAHEQKQAKTKRPREP
 LGSEELPEPGKHPTQDPEAAVSKDTFSYMGESVIVTDCSSNGRKKARAGIG
 VYNGPGHPLNVGILPGRTONRAETHAACKALWQAKONISKVLVLTDSMETINGIT
 NWVQKKNGWRISTGDKVINKEDFELDELTCMDIQMHPHPSGFGVNGEADRLA
 REGAKQSED"
 907..1430
 /gene="Rnh1"

BASE COUNT 384 a 324 c 414 g 308 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.06e-77 Length: 1430
 Score: 1197.50 Matches: 221
 Percent Similarity: 86.71% Conservative: 27
 Best Local Similarity: 77.27% Mismatches: 37
 Query Match: 77.46% Indels: 1
 DB: 10 Gaps: 1

US-10-054-313-1 (1-286) x AF048993 (1-1430)

Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
 Db 49 ATGCGCTGGCTGCTGGCGCTGTCCGCACAGTGCACCTGCCCTGACCCCTGAGCGGA 108
 Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
 Db 109 GGCATTTTGGCGCTGGGATGTTCTATGCGTGCAGGAGGCGCCAGGACCGGAGTCTTC 168
 Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
 Db 169 CTGAGTTGGAGTGTAGTCAAGCCAGGTCGACGGTTCTCTGCTGCCAGTTTAAGAAA 228
 Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
 Db 229 TTTGCCACAGAGATGAGGCTGGGCTTTGTCCAGGAGCTCTTCAAGCCGGATGGTTCA 288
 Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGlnAlaLysProGlyLysArgLeuArg 100
 Db 289 AAGGGCAGGAAGTGCACATGACAGAGTCAAGCGAAGACCAAGCGCGCTCGG 348
 Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
 Db 349 GAGCCCTG---GGTGAAGGGAGAACTTCCAGAGCCAGGGCCAAAGCACACACAGCAG 405
 Qy 121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
 Db 406 GACACCGAGCGCGCTGCTGTAGTGAGCAAGACACATTTTCTTATATGGGAGAGTCACTC 465
 Qy 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160
 Db 466 ATTGCTACACGAGTGGCTGTGCTCCAGTAATGGACGGAAGCGGGCACGAGCAGGAATT 525
 Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLeuArgLeuProGlyArgGln 180
 Db 526 GCGCTTTACTGGGGCCCGAGCCACCTTAAATGTAGCTATAAGGCTTCTCTGGCGCAG 585
 Qy 181 ThrAsnGlnArgAlaGluLeuHisAlaCysLysValLeuGluGlnAlaLysThrGln 200
 Db 586 ACAAACACAGAGCGCGAGATCCATGCGAGCTTGCAGGCCCATCATGAAGCCAAAGGCTCAG 645
 Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
 Db 646 AACATCAGCAGCTGGTTCTGTACACAGACAGCATGTTCAACATCAATGGGATAACTAAC 705
 Qy 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValLys 240
 Db 706 TGGGTTTCAAGGCTCGAAGAAGATGGCTGGAGAACAAAGTACAGGGAAGATGTGTATCAAC 765
 Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
 Db 766 AAGAGGAGTTCATGGAGCTGGAGCAGCTCACTCAGGGCATGGACATCCAGTGCATGCAC 825
 Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGlyGly 280
 Db 826 ATTCTGCTCACTCAGGATTTGTGGCAATTTGTGGCAATTAAGAGGCCGACAGACTGGCAGGAAGGA 885
 Qy 281 AlaLysGlnSerGluAsp 286
 Db 886 GCGAAGCAGTCTGAGGAC 903

RESULT 9
 BC019411
 LOCUS BC019411
 DEFINITION Mus musculus ribonuclease H1, mRNA (cdna clone MGC:30220 IMAGE:5133942), complete cds.
 ACCESSION BC019411
 VERSION BC019411.1 GI:18044435
 KEYWORDS MGC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1409)

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Guntaratne, P.H., Richards, S., Wierley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzyzanski, M.I., Skalska, J., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marr, M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257
12477932
2 (bases 1 to 1409)
Strausberg, R.
Direct Submission
Submitted (19-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Guntaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 41 Row: n Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6755333.

FEATURES
source

1..1409
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:30220 IMAGE:5133942"
/tissue_type="Liver, normal. 5 month old male mouse."
/clone_lib="NCI CGAP_L19"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
1..1409
/gene="Rnaseh1"
/db_xref="LocusID:19819"
/db_xref="MGI:1335073"
17..874
/codon_start=1
/product="ribonuclease H1"
/protein_id="AAH19411.1"
/db_xref="GI:18044436"
/db_xref="LocusID:19819"
/translation="MRLPLSRVTLAVVRLREGICGLGWVAVRGRGTGVFLNSH
ECKQVDRPPARFKFATEDANAFVRSSSSPGSKQSSQKQSVKTSKRPREP

gene

CDS

BASE COUNT 391 a 315 c 400 g 303 t
ORIGIN
LGESELPDPGPKHTRODTEPAAVVSKDTFPMGESVIVTDTGCCSNRKRARAGIG
VTWGGCHPLNIGIRLPGRQTNORABIHAACRAIMQAKAONISKLVLYDSFTINGIT
NNWQKNGKNGWRTSTGKDVINKEDFMELDTQGMIDIONMHIPIHSGFVNEEADRLA
REGAKQSD"

Alignment Scores:
Pred. No.: 2,02e-77 Length: 1409
Score: 1193.50 Matches: 220
Percent Similarity: 86.36% Conservative: 27
Best Local Similarity: 76.92% Mismatches: 38
Query Match: 77.20% Indels: 1
DB: 10 Gaps: 1

US-10-054-313-1 (1-286) x BC019411 (1-1409)

QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 17 ATGCGCTGCTGCTGCGCTGTCCGCGCACAGTGACACTGGCGCTGCTACGCTGAGGCGA 76
QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyValArgGlyThrGlyValPhe 40
Db 77 GCATTTGCGGCTCGGATGTTCTATGCGGTGAGAGAGCCGCGAGGACCGAGTCTTC 136
QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db 137 CTGAGTTGAGTGAGTGCAAGCCAGGCTGGACCGGTTCTCTGCTGCCAGGTTTAAGAA 196
QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 197 TTTCACAGAAAGATGAGCGCTGGGCTTTGTGAGAGCTCTTCAAGCCGCGATGTTCA 256
QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 257 AAAGGCGAGGAAAGTGCATGACAGCAGAGTACAGGTGAGACACAGCAGCGCTCG 316
QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 317 GAGCCCTG---GGTGAAGGGAAGAATTCAGAGCCAGGCGCCAAAGCACACAGCAG 373
QY 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 374 GACACCGACCCGCTGCTGTAGTGACAGGACACGTTTCTTATATGGAGAGTCAGTC 433
QY 141 ValValThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160
Db 434 ATTGCTACAGGATGGCTGTGCTCCAGTAATGACGAGGAGCGGCGACAGCAGGAATT 493
QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLeuArgLeuProGlyArgGln 180
Db 494 GCGTTTACTGGGCGCCAGGCCACCCCTTAATGTAGGTATAGGCTTCTCGGCGACAG 553
QY 181 ThrAsnGlnArgAlaGluLeuHisAlaLysLysAlaLysLysGluValLysGln 200
Db 554 ACAAAACAGAGGCGCGAGATCCATGCGCTGCAAGGCCATCATGCAAGCCAAAGGCTCAG 613
QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyLeuThrAsn 220
Db 614 AACATCAGCAAGCTGGTCTCTGTACACAGACAGCATGTTCCACCATCAATGGGATACTAC 673
QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValLys 240
Db 674 TGGTTTCAGGCTGGAAGAAGAATGGCTGGAGAACAGTACAGGGAAGATGTGATCAAC 733
QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db 734 AAGGAGGACTTCATGGAGCTGGACGAGTCACTCAGGCGCATGCATCCAAATGGATGCAC 793
QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db 794 ATTCTGGTCACTCAGGATTTGTGGCAATGAGAGCGCGACAGACTGGCAGCGGAGGA 853

QY 281 AlalysGlnSerGluasp 286
 |||||
 Db 854 GCGAAGCAGCTCTGAGGAC 871

RESULT 10

AC098850 178965 bp DNA linear PRI 30-DEC-2001
 DEFINITION Homo sapiens chromosome 17, clone RP11-92B11, complete sequence.
 ACCESSION AC098850
 VERSION AC098850.2 GI:17646937
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 178965)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 17, clone RP11-92B11
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 178965)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J.J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Headford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,
 Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
 Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G.,
 Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPeeters,R., Meldrim,J.,
 Meneus,L., Mihova,T., Mienna,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuppback,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thoman,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
 Submitted (03-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 178965)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J.J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Headford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,
 Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
 Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G.,
 Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPeeters,R., Meldrim,J.,
 Meneus,L., Mihova,T., Mienna,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuppback,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thoman,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

JOURNAL
 Submitted (03-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

AUTHORS
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J.J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Headford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,
 Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
 Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G.,
 Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPeeters,R., Meldrim,J.,
 Meneus,L., Mihova,T., Mienna,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuppback,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thoman,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

JOURNAL
 Submitted (30-DEC-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Dec 13, 2001 this sequence version replaced gi:16604041.
 All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RX/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L13829
 Center clone name: 92_B_11

FEATURES

| source | Location/Qualifiers |
|---------------|---|
| unsure | 1..178965 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="17" /map="17" /clone="RP11-92B11" /clone_lib="RPC1-11 Human Male BAC" 987..1002 /note="<30 qual SNGL region" |
| repeat_region | 994..1308 /rpt_family="AluJo" complement(1434..1735) /rpt_family="AluSx" |
| repeat_region | 2075..2095 /rpt_family="AT rich" complement(2205..2509) /rpt_family="AluSx" |
| repeat_region | 2558..2592 /rpt_family="AT rich" 2730..3033 /rpt_family="AluJo" |
| repeat_region | 3034..3339 /rpt_family="AluSx" |
| repeat_region | 4237..4520 /rpt_family="AluSx" |
| repeat_region | 4584..5144 /rpt_family="L1ME" |
| repeat_region | 5180..5481 /rpt_family="AluSx" complement(5496..5542) /rpt_family="AluSx" |
| repeat_region | complement(5543..5678) /rpt_family="AluSx" |
| repeat_region | complement(5801..6120) /rpt_family="AluSx" |
| repeat_region | 6123..6363 /rpt_family="MERS58B" |
| repeat_region | 6423..6533 /rpt_family="L1ME" |
| repeat_region | 6534..6830 /rpt_family="AluSx" |
| repeat_region | 6831..7175 /rpt_family="L1ME" |
| repeat_region | 7310..7435 /rpt_family="L1ME" |
| repeat_region | 7527..7703 /rpt_family="L1ME" |
| repeat_region | 7704..7993 /rpt_family="AluJo" |
| repeat_region | 7994..8309 /rpt_family="L1ME" |
| repeat_region | 8319..8807 /rpt_family="L2" |
| repeat_region | 8893..9575 /rpt_family="L2" |
| repeat_region | 9941..9994 /rpt_family="GA-rich" complement(10332..10491) /rpt_family="MERS3" |
| repeat_region | complement(11158..11392) |

Choepl, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 Dearellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
 Ferreira, P., FitzHugh, M., Forrest, C., Gage, D., Galagan, J.,
 Gadyana, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,
 MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
 McPeeters, R., Meldrim, J., Menus, L., Morrow, J., Naylor, J.,
 Norman, C.H., O'Connor, T., O'Donnell, P., Olivari, T.M., Peterson, K.,
 Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
 Zimmer, A. and Zody, M.

TITLE JOURNAL

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE AUTHORS

3 (bases 1 to 192337)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bada, F., Boguslavsky, L.,
 Bouckgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
 Choepl, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,
 FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,
 Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L.,
 Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., LaRocque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G.,
 MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McKernan, K.,
 McPeeters, R., Meldrim, J., Menus, L., Mihova, T., Mlenga, V.,
 Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,
 O'Donnell, P., O'Neil, D., Olivari, T.M., Oliver, J., Peterson, K.,
 Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riback, M., Riley, R.,
 Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
 Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
 Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
 Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
 Zimmer, A. and Zody, M.

TITLE JOURNAL

Submitted (23-SEP-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Sep 23, 2000 this sequence version replaced gi:9795639.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4191

Center clone name: 219_A15

FEATURES source

Location/Qualifiers

1. .192337
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17"
 /clone="RP11-219A15"
 /clone_lib="RP11-11 Human Male BAC"
 568. 788
 /rpt_family="MLT1C"
 1804. 1957
 /rpt_family="FRAM"
 1964. 2014
 /rpt_family="MIR"
 complement(2029. .2319)
 /rpt_family="LTR16C"
 2463. 2535
 /rpt_family="MIR"

repeat_region
 complement(2614. .2700)
 /rpt_family="MIR3"
 complement(2938. .3185)
 /rpt_family="LIMC/D"
 complement(4802. .4869)
 /rpt_family="L2"
 complement(4871. .5058)
 /rpt_family="LTR16B"
 complement(5150. .5227)
 /rpt_family="LTR16A"
 5353. 5597
 /rpt_family="AluJo"
 5697. 5863
 /rpt_family="MLT1J"
 complement(5909. .6210)
 /rpt_family="AluX"
 complement(6235. .6609)
 /rpt_family="LIMB3"
 6810. 6887
 /rpt_family="AluSg/x"
 complement(6888. .6802)
 /rpt_family="LIMB3"
 7399. 7751
 /rpt_family="MSTD"
 7977. 8102
 /rpt_family="MER5A"
 complement(8265. .8596)
 /rpt_family="LIME3"
 complement(10173. .10360)
 /rpt_family="MIR"
 complement(10370. .10452)
 /rpt_family="MIR"
 10648. 11010
 /rpt_family="LIME4A"
 complement(11168. .11273)
 /rpt_family="LIMC4"
 complement(11274. .11583)
 /rpt_family="AluSx"
 complement(11584. .11603)
 /rpt_family="LIMC4"
 12176. 12289
 /rpt_family="MIR"
 complement(12402. .12464)
 /rpt_family="L2"
 13177. 13351
 /rpt_family="MER69"
 13950. 14083
 /rpt_family="MIR"
 complement(14660. .15282)
 /rpt_family="LIMC4a"
 complement(15834. .15968)
 /rpt_family="MIR3"
 complement(16176. .16237)
 /rpt_family="L2"
 17027. 17098
 /rpt_family="MIR"
 complement(17196. .17328)
 /rpt_family="L2"
 complement(17390. .17492)
 /rpt_family="L2"
 18087. 18429
 /rpt_family="L2"
 18524. 18675
 /rpt_family="MER3"
 18848. 19158
 /rpt_family="AluSp"
 complement(19554. .19800)
 /rpt_family="AluSx"
 complement(20989. .21284)
 /rpt_family="AluY"
 21286. 21365
 /rpt_family="MER3"
 21393. 21576

```

repeat_region      /rpt family="L2"
21627..21701
/rpt family="MIR"
21770..22001
/rpt family="LTR14A"
complement(22014..22401)
/rpt family="L1M4A4"
23317..23339
/rpt family="AT rich"
complement(23367..23502)
/rpt family="MIR"
24118..24141
/rpt family="TCTG)n"
complement(24473..24616)
/rpt family="MIR3"
complement(25652..26053)
/rpt family="L2"
28293..28376
/rpt family="MIR3"
34302..34408
/rpt family="MIR"
complement(35693..35781)
/rpt family="MIR"
37220..37405
/rpt family="L2"
complement(37431..37554)
/rpt family="MIR"
38931..39908
/rpt family="LTRS"
40343..40448
/rpt family="MIR"
40552..40629
/rpt family="MIR"
complement(41149..41275)
/rpt family="MIR"
complement(41355..41506)
/rpt family="MIR3"

Alignment Scores:
Pred. No.:      1.22e-73      Length:      192337
Scores:         1174.00      Matches:      236
Percent Similarity: 88.36%      Conservative: 7
Best Local Similarity: 85.82%      Mismatches: 18
Query Match:     75.94%      Indels:      16
DB:              9          Gaps:        3

US-10-054-313-1 (1-286) x ACO22596 (1-192337)

Qy      1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArg 20
Db      191561 ATGAGCGCGCTTCTGTCCCTGGCGGC-AGAGTCGCGCTTGGCGCGCTTGCCTTGCCTGCGCGCGC 191503

Qy      21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyAsGlyThrGlyValPhe 40
Db      191502 ---TCTCGGGTTCGGGATGTTCTATGCCGTGAGAGGGCGGCAAGTCCGCGGTCTTT 191446

Qy      41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheGlyLys 60
Db      191445 CTGACCGGGAATGAGTGCAGAACACAGAGTGCAGCGGTTTCTGCTGCGCAGATTTAAGAAG 191386

Qy      61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaAspProGluValSer 80
Db      191385 TTTCACAGAGGATGAGCGCTGGGACTTTTTCAGAAATCTGACAGCCCGGAAGTTTCA 191326

Qy      81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db      191325 GAAGGCGAGGAAATCAACATCGGACAGAAAGCGAGCAAGCGACTCCGT 191266

Qy      101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMet----- 118
Db      191265 GAGCCATCGATGGAGATGGAGATGAAAGCGCAGACCATATGC-AAAGCATGAAAGCG 191207

Qy      119 -----LysProSerValGluProAlaProPrcValSerArgAspThr 132

```

```

Db      191206 AGCGTGAAGCCGAGCGTGAAGCCGAGCGTGAAGCCGCGCTCCAGTTAGAGACACAG 191147
Qy      133 PheSerTyrMetGlyAspPheValValTyrThrAspGlyCysCysSerSerAsnGly 152
Db      191146 TTTTCTACATGGGAGACTTGTGCTGTCTACGCTGATGGCTGTCTCCAGTAATGGG 191087
Qy      153 ArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGlyHisProLeuAsnVal 172
Db      191086 CGTAGAAAGCCACGAGCAGGAATCCGCGTTTACTGGGCGCGGGTATCCCTTTAACTGTA 191027
Qy      173 GlyLeuArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLys 192
Db      191026 GGCATTAGACTTCTTGGGTGGCAGACAAACCAAGAGCGGAAATTCATGCAGCTGCAAG 190967
Qy      193 AlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMet 212
Db      190966 GCCATTGAACAAGCACAGACTCAGAAATCAATAAATCGTTCTGTATACAGACAGTATG 190907
Qy      213 PheThrIleAsnGlyLeuThrAsnTrpValGlnGlyTyrLysLysAsnGlyTyrLysThr 232
Db      190906 TTTACGATAAATGGTATTAACCTGAGTTTTCAGGTTGGAAGAAAATATGGTGGAGACA 190847
Qy      233 SerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGln 252
Db      190846 AGTGACAGGAAAGAGTGTATCAACAAGAGGACTTCGTGGCACTGGAGAGGCTGACCCAG 190787
Qy      253 GlyMetAspIleGlnTrp-----MetHisValPro 262
Db      190786 GGGATGGACATTCACTGGGCGAGTATCTTAATGTTTCATGCCCCA 190742

RESULT 12
AC126352/c      211403 bp      DNA      linear      PRI 19-MAR-2003
LOCUS      Homo sapiens chromosome 17, clone CTD-2509K7, complete sequence.
DEFINITION      AC126352
ACCESSION      AC126352.6 GI:29124211
VERSION      HTG.
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 211403)
AUTHORS      Birren,B., Nusbaum,C. and Lander,E.
TITLE      Homo sapiens chromosome 17, clone CTD-2509K7
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 211403)
AUTHORS      Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barra,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Kartas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,X.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Melchrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
TITLE      Direct Submission
JOURNAL      Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE      3 (bases 1 to 211403)
AUTHORS      Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barra,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,

```

Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meidrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (20-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 211403)

Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Levine, R., Kanat, A., Karatas, A., Kells, C., Landers, T., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meidrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (19-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 19, 2003 this sequence version replaced gi:25140067.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L27684

Center clone name: 2509_K7

FEATURES

source

Location/Qualifiers
1. .211403
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="CTD-2509K7"
/clone_lib="CITD2 Human BAC"
3. .67
/rpt_family="MER5A"
complement(230..561)
/rpt_family="L1ME3"
1679..1776
/rpt_family="L3b"
complement(2138..2325)
/rpt_family="MIR"
complement(2335..2417)
/rpt_family="MIR"

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region
2613..2966
/rpt_family="L1ME4a"
complement(3133..3237)
/rpt_family="L1MC4"
repeat_region
complement(3239..3547)
/rpt_family="AlusX"
4140..4268
/rpt_family="MIR"
complement(4365..4427)
/rpt_family="L2"
5140..5314
/rpt_family="MER69"
5913..6046
/rpt_family="MIR"
complement(6623..7245)
/rpt_family="L1MC4a"
complement(7797..7931)
/rpt_family="MIR3"
complement(8139..8200)
/rpt_family="L2"
8990..9061
/rpt_family="MIR"
complement(9159..9291)
/rpt_family="L2"
complement(9353..9455)
/rpt_family="L2"
complement(9570..9638)
/rpt_family="L3b"
10050..10392
/rpt_family="L2"
10487..10638
/rpt_family="MER3"
10811..11121
/rpt_family="AluSp"
complement(11517..11763)
/rpt_family="AluX"
11534..11544
/note="<30 qual SINGL region"
11554..11592
/note="<30 qual SINGL region"
complement(12952..13247)
/rpt_family="AluY"
13249..13328
/rpt_family="MER3"
13356..13539
/rpt_family="L2"
13590..13664
/rpt_family="MIR"
13733..13964
/rpt_family="L1R14A"
complement(13965..14364)
/rpt_family="L1PB4"
15280..15302
/rpt_family="AT-rich"
complement(15330..15465)
/rpt_family="MIR"
16082..16105
/rpt_family="(TCTG)n"
complement(16437..16580)
/rpt_family="MIR3"
complement(17671..18017)
/rpt_family="L2"
complement(18266..18369)
/rpt_family="L3b"
20257..20340
/rpt_family="MIR3"
26262..26368
/rpt_family="MIR"
complement(27651..27739)
/rpt_family="MIR"
29178..29363
/rpt_family="L2"
complement(29389..29512)

```

/rpt family="MIR"
30886..31865
/rpt family="LTRS_Hs"
32286..32405
/rpt family="MIR"
32509..32586
/rpt family="MIR"
complement(33106..33232)
/rpt family="MIR"
complement(33312..33463)
/rpt family="MIR3"
34251..34325
/rpt family="(TGG)n"
complement(35953..36046)

repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region

Alignment Scores:
Pred. No.: 6,066-73 Length: 211403
Score: 1165.00 Matches: 235
Percent Similarity: 88.00% Conservative: 7
Best Local Similarity: 85.45% Mismatches: 19
Query Match: 75.36% Indels: 16
DB: 9 Gaps: 3

US-10-054-313-1 (1-286) x AC126352 (1-211403)
Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 183462 ATGACCGCGCTCTGCTCCCTGCGCGC-AGAGTCGCTTGGCGCGCTTGGCGCGCGC 183404
Qy 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgArgGlyArgGlyThrGlyValPhe 40
Db 183403 ---TCTCGCGGTTCCGGATGTTCTATGCGGTGAGAGGCGCGCAAGTCCGCGCTCTTT 183347
Qy 41 LeuTrpTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db 183346 CTGACCGGAATGAGTGAACACACAGGTGGACCGGTTTCTGCTGCCAGATTTAAGAG 183287
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 183286 TTTGCCACAGAGATGAGCGCTTGGGCTTTGTTCAGGAAATCTGCAAGCCGCGAAGTTTCA 183227
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLysArg 100
Db 183226 GAAGCGCAGAAATCAACATGGACAGAAATCGGAGACGAAGCCAGCAAGCGACTCCGT 183167
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMet----- 118
Db 183166 GAGCCACTGGATGGAGATGGAGATGAAGCGCAGAGCCATATGC-AAGCAGATGAAGCGC 183108
Qy 119 -----LysProSerValGluProAlaProProValSerArgAspThr 132
Db 183107 AGTGTGAAGCCGAGCGTGAAGCGCGAGCGTGAAGCGCGCGCTCCAGTTAGCAGAGACG 183048
Qy 133 PheSerTyrMetGlyAspPheValValValTyrThrAspGlyCysCysSerSerAsnGly 152
Db 183047 TTTTCTACATGGGAGACTTCGTCGTCGTCACGCTGATGCTGCTGCTCCAGTAATGG 182988
Qy 153 ArgArgLysProArgAlaGlyLeGlyValTyrTrpGlyProGlyHisProLeuAsnVal 172
Db 182987 CGTAGAAGCCAGCAGCAAGAAATCGCGGTTTACTGGGGCGCGGCTATCCTTTAACTGTA 182928
Qy 173 GlyValArgLeuProGlyArgGlnThrAsnGlnArgAlaGluLeuHisAlaAlaCysLys 192
Db 182927 GGCATTAGACTTCTGGGTGGCAGACATCCAAAGAGCGGAATTCATGCGAGCTGCAAG 182868
Qy 193 AlaLeuGluGlnAlaLysThrGlnAsnLeuValValLeuValLeuTyrThrAspSerMet 212
Db 182867 GCCATTGAACAGCAGACTCAGAAAAATCAATAAATCGTTCTGTATACAGACAGTATG 182808
Qy 213 PheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThr 232
Db 182807 TTTACGATAAATGGTATAACTAACTACGTTCAAGGTTTGAAGAAAAATGGGTGAAGACA 182748

```

```

Qy 233 SerAlaGlyCysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGln 252
Db 182747 AGTGCAGGAAAGAGGTGATCAACAAGAGGAGCTTCGTGGCAGTGGAGAGGCTGACCCAG 182688
Qy 253 GlyMetAspIleGlnTrp-----MetHisValPro 262
Db 182687 GCGATGGACATTCAGTGGCGAGATATCTTAATGTTTCATGTCCCA 182643

RESULT 13
AC090774 207418 bp DNA linear PRI 28-JUN-2002
LOCUS Homo sapiens chromosome 17, clone RP11-344E13, complete sequence.
DEFINITION AC090774
VERSION AC090774.6 GI:21617728
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 207418)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone RP11-344E13
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 207418)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
Camrata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeAtellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Glade,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hartford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldrum,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,E., Roman,J., Rosett,M.,
Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Sounez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (10-MAR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 207418)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camrata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cooke,A., Cooke,P., DeAtellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (27-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 207418)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,

```

Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collamore, A., Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Farq, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Melidrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (28-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 27, 2002 this sequence version replaced gi:21426207.
All repeats were identified using RepeatMasker.
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: Li2729
Center clone name: 344_E_13

FEATURES

Source

Location/Qualifiers
1. .207418
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="17"
/map="17"
/clone="RP11-344E13"
/clone_lib="RP11-11 Human Male BAC"

repeat_region
1. .264
/rpt_family="LTR35"
repeat_region
285. .499
/rpt_family="LIPA13"
repeat_region
949. .1259
/rpt_family="AluY"
repeat_region
1268. .1401
/rpt_family="L1MC/D"
repeat_region
2066. .2204
/rpt_family="LTR67"
repeat_region
2210. .2340
/rpt_family="L1MC/D"
repeat_region
complement(2412. .2546)
/rpt_family="L2"
repeat_region
complement(3163. .3459)
/rpt_family="AluSg"
repeat_region
6492. .6534
/rpt_family="TG)n"
repeat_region
6753. .6808
/rpt_family="AT rich"
repeat_region
complement(7108. .7407)
/rpt_family="AluY"
repeat_region
7825. .8108
/rpt_family="AluY"
repeat_region
complement(8512. .8547)
/rpt_family="MADE1"
repeat_region
8875. .8940
/rpt_family="L2"
repeat_region
9921. .10228
/rpt_family="AluSx"
repeat_region
10713. .11010
/rpt_family="AluJb"

repeat_region
12697. .12804
/rpt_family="FLAM_A"
complement(12850. .13549)
/rpt_family="LIPA10"
repeat_region
13550. .13850
/rpt_family="AluSp"
complement(13851. .14554)
/rpt_family="LIPA10"
repeat_region
14580. .15680
/rpt_family="LIPA15"
repeat_region
15862. .15967
/rpt_family="L2"
repeat_region
15968. .15988
/rpt_family="TGAA)n"
repeat_region
16061. .17124
/rpt_family="L1M4"
repeat_region
17590. .17664
/rpt_family="L1M4"
repeat_region
17665. .17967
/rpt_family="AluY"
repeat_region
17968. .17979
/rpt_family="L1M4"
repeat_region
17980. .18006
/rpt_family="(CAAA)n"
repeat_region
18007. .18381
/rpt_family="L1M4"
repeat_region
20417. .20580
/rpt_family="AluY"
repeat_region
20586. .21111
/rpt_family="L1M4"
repeat_region
complement(21112. .21403)
/rpt_family="AluSx"
repeat_region
21404. .22015
/rpt_family="L1M4"
repeat_region
22063. .22094
/rpt_family="AT rich"
repeat_region
23322. .23380
/rpt_family="(TA)n"
complement(23381. .23670)
/rpt_family="AluSg"
repeat_region
complement(25468. .26047)
/rpt_family="L1ME3A"
repeat_region
complement(26143. .26832)
/rpt_family="L1ME3A"
repeat_region
27610. .27727
/rpt_family="AluY"
repeat_region
27728. .28199
/rpt_family="L1PA4"
repeat_region
28200. .28408
/rpt_family="AluY"
repeat_region
28936. .29315
/rpt_family="MSTD"
complement(29697. .30168)
/rpt_family="L1MC2"
repeat_region
30281. .30372
/rpt_family="(TA)n"
complement(30403. .30510)
/rpt_family="MIR"
repeat_region
30962. .31259
/rpt_family="MER44D"
repeat_region
complement(31773. .32400)
/rpt_family="L2"
repeat_region
32401. .32428
/rpt_family="AT rich"

Alignment Scores:

Fred. NO.: 6.61e-72 Length: 207418
Score: 1150.50 Matches: 234
Percent Similarity: 89.89% Conservative: 6
Best Local Similarity: 87.64% Mismatches: 20
Query Match: 74.42% Indels: 10
DB: 9 Gaps: 3

US-10-054-313-1 (1-286) x AC090774 (1-207418)

| | | | |
|----|-------|--|-------|
| Qy | 1 | MetSerTrpIeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg | 20 |
| Db | 58364 | ATGAGCGGGCTTCTGTCTCTGGCCGC-AGAGTCGCTTGGCCGCCTTCCCTGC---CGC | 58419 |
| Qy | 21 | GlySerArgGlyPheGlyMetPheTyrAlaValArgAtsGlyArgLysThrGlyValPhe | 40 |
| Db | 58420 | TGCTCTCACGGGTTCCGGATTGTTCTATGCCGTGAGAGGGGGCGCAAGCCGGGGTCTTT | 58479 |
| Qy | 41 | LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys | 60 |
| Db | 58480 | CTGACCGGGAATGAGTCGCAAGACACAGGTGGACCGGTTTCTCTGCTGCCAGATTTAAGAAG | 58539 |
| Qy | 61 | PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer | 80 |
| Db | 58540 | TTTGCCACAGAGATGAGACCTGGGACTTTGTGAGGAATCTGCAAGCCCGGAAGTTTCA | 58599 |
| Qy | 81 | GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg | 100 |
| Db | 58600 | GAAGGCGAGAAATCAACATGGACAGAATCAGAGACGAAAGCCAGCAAGCACTCCGT | 58659 |
| Qy | 101 | GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro | 120 |
| Db | 58660 | GAGCCATGATGATGAGATGGAGATGAAGCGCAGAGCCATATGC-AAGCACATGAAGCGC | 58718 |
| Qy | 121 | SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal | 140 |
| Db | 58719 | AGGGTAGGCGCGCGCTCCAGTTACAGAGACACGTTTCTACATGGGAGACTTCGTC | 58778 |
| Qy | 141 | ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle | 160 |
| Db | 58779 | GTCGCTACGCTGATGGCTGCTGCTCCAGTAAATGGCGGTAGAAAGCCAGCAGCAATC | 58838 |
| Qy | 161 | GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln | 180 |
| Db | 58839 | CGCGTTTACTTGGGGCGGGCTATCTCTTAAATGTAGGCATTAGACTTCTCTGGGTGGCAG | 58898 |
| Qy | 181 | ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluAlaLysThrGln | 200 |
| Db | 58899 | ACAACCAAGAGCGGAAATTCATGCAGCTACAGGCCATTGACACAGCAAGAACTCAA | 58958 |
| Qy | 201 | AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn | 220 |
| Db | 58959 | AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTGT---AAC | 59014 |
| Qy | 221 | TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn | 240 |
| Db | 59015 | TGGGTTCAAGGTTGGAGAAATAATGATGGAACAAAGTGCAGGGGAAAGAGGTGATCAAC | 59074 |
| Qy | 241 | LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrp----- | 258 |
| Db | 59075 | AAAGAGACCTTCGTGGCACTGGAGAGGCTGACCCAGGGGATGGACATTCAGTGGCGAGT | 59134 |
| Qy | 259 | -----MethisValPro 262 | |
| Db | 59135 | ATCTTAATATGTTATGTCCTCCA 59155 | |

| | |
|------------|---|
| RESULT | 14 |
| AC107926 | |
| LOCUS | Homo sapiens, clone RP11-381P6, complete sequence. |
| DEFINITION | AC107926 linear |
| ACCESSION | AC107926 |
| VERSION | AC107926.6 GI:21637512 |
| KEYWORDS | HTG. |
| SOURCE | Homo sapiens (human) |
| ORGANISM | Homo sapiens |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| REFERENCE | 1 (bases 1 to 176040) |
| AUTHORS | Birren,B., Nusbaum,C. and Lander,E. |
| TITLE | Homo sapiens, Clone RP11-381P6 |
| JOURNAL | Unpublished |

REFERENCE
AUTHORS

2 (bases 1 to 176040)
 Birren-Bern, Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
 Chepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
 Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R.,
 Landers, T., Lechoczky, J., Levine, R., Liu, G., Maclean, C.,
 MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
 McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,
 Mlewa, V., Murphy, T., Navlor, J., Nguyen, C., Nicol, R., Norbu, C.,
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Petersen, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
 Retta, R., Riback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuppach, R., Seaman, S.,
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 176040)

TITLE
JOURNAL

Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 176040)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Brown, A., Cantarat, J., Campopiano, A., Chang, J., Chararo, B., Choepsey, X., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeAllencan, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Lander, T., Lechoczky, J., Levine, R., Lindblad, T., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Menseis, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuppback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zaincun, J., Zembek, J., Zimmer, A., and Zody, M.

TITLE
JOURNAL

Direct Submission
Submitted (15-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 176040)
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collumors, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, U., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karakas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Miengwa, V., Murphy, I., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomas, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (01-JUL-2002) Whitehead Institute/MIT Center for Genome

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 1, 2002 this sequence version replaced gi:21427785.
 All repeats were identified using RepeatMasker:
 Smit, A.P.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L21134
 Center clone name: 381_P_6

Only the first 176.0 kb of this clone are being submitted.
 The remainder overlaps accession number AC090774 [WIGOR project
 L12729].

FEATURES

| | | | |
|---------------|-------------------------------------|--|--|
| source | Location/Qualifiers | | |
| repeat_region | 1. .176040 | | |
| repeat_region | /organism="Homo sapiens" | | |
| repeat_region | /mol_type="genomic DNA" | | |
| repeat_region | /db_xref="taxon:9606" | | |
| repeat_region | /clone="RP11-381P6" | | |
| repeat_region | /clone_lib="RP11-11 Human Male BAC" | | |
| repeat_region | 7. .233 | | |
| repeat_region | /rpt_family="AluJo" | | |
| repeat_region | 477. .662 | | |
| repeat_region | /rpt_family="HAL1" | | |
| repeat_region | 1499. .1785 | | |
| repeat_region | /rpt_family="AluSc" | | |
| repeat_region | 1793. .1850 | | |
| repeat_region | /rpt_family="L1MB5" | | |
| repeat_region | 1851. .2131 | | |
| repeat_region | /rpt_family="AluY" | | |
| repeat_region | 2135. .2163 | | |
| repeat_region | /rpt_family="(CAA)n" | | |
| repeat_region | 2164. .2936 | | |
| repeat_region | /rpt_family="L1MB5" | | |
| repeat_region | 3425. .3575 | | |
| repeat_region | /rpt_family="MIR" | | |
| repeat_region | 4063. .4183 | | |
| repeat_region | /rpt_family="GA-rich" | | |
| repeat_region | 4740. .4778 | | |
| repeat_region | /rpt_family="AT rich" | | |
| repeat_region | complement(4785..4940) | | |
| repeat_region | /rpt_family="FRAM" | | |
| repeat_region | 5727. .6082 | | |
| repeat_region | /rpt_family="L1MC4a" | | |
| repeat_region | 6096. .6306 | | |
| repeat_region | /rpt_family="L1MC4a" | | |
| repeat_region | 6447. .7402 | | |
| repeat_region | /rpt_family="L1MB7" | | |
| repeat_region | 7403. .7706 | | |
| repeat_region | /rpt_family="AluSc" | | |
| repeat_region | 7707. .7750 | | |
| repeat_region | /rpt_family="L1MB7" | | |
| repeat_region | 7751. .8105 | | |
| repeat_region | /rpt_family="AluY" | | |
| repeat_region | 8106. .8680 | | |
| repeat_region | /rpt_family="L1MB7" | | |
| repeat_region | 9419. .9524 | | |
| repeat_region | /rpt_family="MER103" | | |
| repeat_region | 10074. .10358 | | |
| repeat_region | /rpt_family="L1MC5" | | |
| repeat_region | 11023. .11208 | | |
| repeat_region | /rpt_family="L2" | | |
| repeat_region | 11223. .11436 | | |
| repeat_region | /rpt_family="YIR" | | |
| repeat_region | 12151. .12112 | | |
| repeat_region | /rpt_family="(CA)n" | | |
| repeat_region | complement(13126..13321) | | |
| repeat_region | /rpt_family="MLT1E2" | | |

| | | | |
|---------------|--------------------------|--|--|
| repeat_region | 13920. .14143 | | |
| repeat_region | /rpt_family="L2" | | |
| repeat_region | 14198. .15784 | | |
| repeat_region | /rpt_family="L1PA16" | | |
| repeat_region | 15787. .15823 | | |
| repeat_region | /rpt_family="(TC)n" | | |
| repeat_region | complement(16696..16840) | | |
| repeat_region | /rpt_family="FRAM" | | |
| repeat_region | complement(17030..17339) | | |
| repeat_region | /rpt_family="AluSc" | | |
| repeat_region | 17575. .17802 | | |
| repeat_region | /rpt_family="L1ME2" | | |
| repeat_region | 17839. .17938 | | |
| repeat_region | /rpt_family="L1ME2" | | |
| repeat_region | 17939. .18247 | | |
| repeat_region | /rpt_family="AluSc" | | |
| repeat_region | 18248. .18793 | | |
| repeat_region | /rpt_family="L1ME2" | | |
| repeat_region | 19076. .19398 | | |
| repeat_region | /rpt_family="LTR35" | | |
| repeat_region | 19399. .19633 | | |
| repeat_region | /rpt_family="L1PA13" | | |
| repeat_region | 20083. .20393 | | |
| repeat_region | /rpt_family="AluY" | | |
| repeat_region | 20402. .20535 | | |
| repeat_region | /rpt_family="L1MC/D" | | |
| repeat_region | 21200. .21338 | | |
| repeat_region | /rpt_family="LTR67" | | |
| repeat_region | 21344. .21474 | | |
| repeat_region | /rpt_family="L1MC/D" | | |
| repeat_region | complement(21546..21680) | | |
| repeat_region | /rpt_family="L2" | | |
| repeat_region | complement(22297..22593) | | |
| repeat_region | /rpt_family="AluSc" | | |
| repeat_region | 25826. .25660 | | |
| repeat_region | /rpt_family="(TG)n" | | |
| repeat_region | 25879. .25934 | | |
| repeat_region | /rpt_family="AT rich" | | |
| repeat_region | complement(26234..26533) | | |
| repeat_region | /rpt_family="AluY" | | |
| repeat_region | 26951. .27234 | | |

| | | | |
|------------------------|----------|---------------|--------|
| Alignment Scores: | | | |
| Pred. No.: | 1.26e-71 | Length: | 176040 |
| Score: | 1145.50 | Matches: | 233 |
| Percent Similarity: | 89.51% | Conservative: | 6 |
| Best Local Similarity: | 87.27% | Mismatches: | 21 |
| Query Match: | 74.09% | Indels: | 10 |
| DB: | 9 | Gaps: | 3 |

| | |
|---|--|
| US-10-054-313-1 (1-286) x AC107926 (1-176040) | |
| QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20 | |
| DB 77485 ATGAGCGGCTTCTGTTCTGCGCGC-AGAGTCGCTTGGCGCTTGGCGCTGCGC-CCG 77540 | |
| QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgLysThrGlyValPhe 40 | |
| DB 77541 TGTCTCACGGGTTCCGGATGTTCTATCCGTGAGAGGGCGGCAAGACCGGGGCTTTT 77600 | |
| QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60 | |
| DB 77601 CTGACCGGGAATGAGTGCAGAGCACAGTGGACCGGTTTCTGCTGCAGATTTAAGAAG 77660 | |
| QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80 | |
| DB 77661 TTTGCCACAGAGATGAGACCTGGGACTTTGTCAGGAATCTGCAAGCCCGAATTTC 77720 | |
| QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100 | |
| DB 77721 GAAGGCGAGGAATATCAATGACATGGACAGATCAGACGAGGAAGCCACCAAGCGACTCGT 77780 | |
| QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120 | |


```

Db 77781 GAGCCACTGGATGAGATGAAAGGCGACGATATGC-AAGCACATCAAGCCG 77839
Qy 121 SerValGluProAlaProValSerArgAspThrPheSerTyMetGlyAspPheVal 140
Db 77840 AGGTGAAGCGCGCTCCAGTTAGCAGACACGTTTCCACATGGGAGACTTCGTC 77899
Qy 141 ValValTyThrAspGlyCySeSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
Db 77900 GTCGTCACGGCTGATGCTGCTCCAGTAATGGCGTAGAAGCCACGAGGAGATC 77959
Qy 161 GlyValTyTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db 77960 CGCGTTTACTGGGGCGGGCTATCCCTTTAATATAGGCATTAGACTTCTGGGTGGCAG 78019
Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAalaCyLysAlaIleGluGlnAlaLysThrGln 200
Db 78020 ACNAAACCAAGAGCGGAATTCATGCAGCTTACAAGCCATTGAACCAAGCAAGACTCAA 78079
Qy 201 AsnIleAsnLysLeuValLeuTyThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 78080 AACATCAATAAAGTGTCTGTATACAGACAGTAGTATGTTACGATAAATGGTGT----AAC 78135
Qy 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
Db 78136 TGGGTTCAAGGTTGGAAGAAAATGGATGGAGACAGTGCAGGGAAGAGGTGACCAAC 78195
Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrp----- 258
Db 78196 AAAGAGGACTTCGTGGCAGCTGAGAGGCTGACCCAGGGGATGGACATTTCAGTGGCGAGT 78255
Qy 259 -----MetHisValPro 262
Db 78256 ATCTTAATGTTCAATGTCCTCCA 78276

```

RESULT 15

HS667H12
LOCUS Human DNA sequence from clone RP4-667H12 on chromosome 1q32.1-41,
complete sequence.

ACCESSION AL035414.30 GI:6822166

VERSION AL035414.30 GI:6822166

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 131239)

Ellington,A.

Direct Submission

Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Jan 30, 2000 this sequence version replaced gi:6782338.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em, ENBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping

Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr1

----- Genome Center

Center: Wellcome Trust Sanger Institute

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

RP4-667H12 is from the library RPCI-4 constructed by the group of Pierer de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: PCIPAC2.

FEATURES

Location/Qualifiers
source
1..131239
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="q32.1-41"
/clone="RP4-667H12"
/clone_lib="RPCI-4"

BASE COUNT 35649 a 28535 c 29022 g 38033 t
ORIGIN

Alignment Scores:

Fred. No.: 2.04e-62 Length: 131239
Score: 1016.00 Matches: 209
Percent Similarity: 79.37% Conservative: 18
Best Local Similarity: 73.08% Mismatches: 58
Query Match: 65.72% Indels: 2
DB: 9 Gaps: 1

US-10-054-313-1 (1-286) x HS667H12 (1-131239)

Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 85425 ATGTGCCAGCTTCTGTCCCTGCGCCACACAGATCACCTTGATAGCTGTGGCTGCCACAG 85484
Qy 21 GlySerArgGlyPheGlyMetPheTyAlaValArgGlyArgLysThrGlyValPhe 40
Db 85485 GACTGCTGCGAGCTCAGGATGCTATGCTGTGGGGAGGCGCTGCAGGACTGGGTCCTT 85544
Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db 85545 CTGACCTCGAATGAATATAGACATAGTGGACCTATTTCTGCTGCCAGATTTAAGAAG 85604
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 85605 TTTGCCACAGAAATAGGCTGGGCTTTGTCAAGAAATCTGCAAGCCACAGAGTTCA 85664
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 85665 GAAGGCGCAAAATAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 85724
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyAlaLysHisMetLysPro 120
Db 85725 GAGCCACTGAATGAGATGGAGATGAAGTGAAGTCTTGTGCAAGAGCTCTGTGAGCAAA 85784
Qy 121 SerValGluProAlaProValSerArgAspThrPheSerTyMetGlyAspPheVal 140
Db 85785 ATCATAGACCGGCACCTCGATGAGCAGACACA--TTTTCTTACATCGAGGCTTTGTC 85842
Qy 141 ValValTyThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
Db 85843 GTCGTCACACTGATGGCTACTGCTCCAGTAATGGGTGGAGGAGCACTAGTGGGAAC 85902
Qy 161 GlyValTyTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db 85903 GACGTTTACTGGGGCGCGGCTATCCCTTTAATATAGGATTAGACTTCTCTGGGTACAG 85962
Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAalaCysLysAlaIleGluGlnAlaLysThrGln 200

```

Db      85963  ACAAACCAAGAGCAGAAAATCATGCGCTGCACACACATGAAACAGCAAGGCTGAA 86022
Qy      201  AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db      86023  AACATCAATAAATTGGTCTGTACACAGATAGTATGTTTGTATAAATGGTATAACTAAC 86082
Qy      221  TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
Db      86083  TGGGTTCAAGGTAGGAAGAAGAACACAGGTGGAAGACAGATACAGGGAATAGGTGATGAAC 86142
Qy      241  LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db      86143  AAAAAGGACTTGTGGCCTGGAGAGCTCAGTCAGGCATGGACATTCAGTGGATACTT 86202
Qy      261  ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db      86203  GTTCCTGGGTCAATCCAGGGTGTATTGGCAATGAAGAAGCTGATAGATTAGCAAGA---GGA 86259
Qy      281  AlaLysGlnSerGluAsp 286
Db      86260  GCTAAACATTCGAAAGAC 86277

```

Search completed: December 17, 2003, 20:35:04
 Job time : 4343 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 17, 2003, 18:02:28 : Search time 318 seconds

(without alignments)
2427.796 Million cell updates/sec

Title: US-10-054-313-1

Perfect score: 1546

Sequence: 1 MSWLLFLAHRVALAALPCRR.....FIGNEADRLAREGAKQSED 286

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-O=/cgn2.1/USPTO.spool/US10054313/runat.17122003.150742.24169/app.query.fasta_1.455
-DB=Geneseq.15Jun03 -OPMT=fastap -SUPFIX=ring -MINMATCH=0.1 -LOOPFCI=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054313 @CEN.1.1.312 @runat.17122003.150742.24169 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq.19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|---------|----|-----------|--------------------|
| 1 | 1525 | 98.7 | 2129 | 23 | ABV30263 | Human prostate exp |
| 2 | 1523 | 98.5 | 1150 | 21 | AA251265 | Human RNA associat |
| 3 | 905 | 58.5 | 545 | 23 | ABV45787 | Human prostate exp |
| 4 | 666.5 | 43.1 | 691 | 23 | ABV15988 | Human prostate exp |
| 5 | 508.5 | 32.9 | 389 | 21 | AAH30241 | Human colon cancer |
| 6 | 434 | 28.1 | 473 | 25 | ABX41815 | Bovine EST associ |
| 7 | 422 | 27.3 | 1071 | 23 | ABL03247 | Drosophila melanog |
| 8 | 399 | 25.8 | 407 | 25 | ABX38735 | Bovine EST associ |
| 9 | 364 | 23.5 | 764 | 24 | ABX77198 | Frog embryonic gen |
| 10 | 356.5 | 23.1 | 3290 | 23 | ABL03246 | Drosophila melanog |
| c 11 | 272 | 17.6 | 11921 | 22 | AAF28553 | Genomic fragment # |
| 12 | 244 | 15.8 | 378 | 16 | AAI24248 | Human gene signatu |
| c 13 | 239 | 15.5 | 7874 | 20 | AAI20568 | Polynucleotide seq |
| 14 | 236.5 | 15.3 | 462 | 22 | AAI20568 | Haemophilus influe |
| 15 | 236.5 | 15.3 | 462 | 24 | ABK64944 | DNA encoding Haemo |
| 16 | 236.5 | 15.3 | 1001 | 22 | AAI21449 | Haemophilus influe |
| 17 | 236.5 | 15.3 | 1001 | 24 | ABK37829 | DNA sequence upstr |
| c 18 | 236.5 | 15.3 | 1830121 | 17 | AAI42063 | Haemophilus influe |
| 19 | 235.5 | 15.2 | 2048 | 20 | AAI98244 | Contig 1344 identi |
| 20 | 234.5 | 15.2 | 435 | 25 | ABZ40958 | N. gonorrhoeae nuc |
| 21 | 229.5 | 14.8 | 36471 | 21 | AAI81453 | N. meningitidis pa |
| 22 | 229.5 | 14.8 | 349980 | 21 | AAI21611 | Neisseria meningit |
| 23 | 229.5 | 14.8 | 837096 | 21 | AAI81489 | N. meningitidis pa |
| 24 | 221.5 | 14.3 | 801 | 20 | AAI98246 | Nucleotide sequenc |
| 25 | 213.5 | 13.8 | 465 | 16 | AAQ96268 | E.coli ribonucleas |
| 26 | 211.5 | 13.7 | 465 | 16 | AAQ96267 | E.coli ribonucleas |
| 27 | 209.5 | 13.6 | 757 | 12 | AAQ12780 | Encodes stabilised |
| c 28 | 203 | 13.1 | 349980 | 24 | ABQ1849 | Bifidobacterium lo |
| 29 | 187 | 12.1 | 32502 | 23 | AAI59542 | Propionibacterium |
| 30 | 182.5 | 11.8 | 6397 | 23 | ABZ03444 | Drosophila melanog |
| 31 | 176.5 | 11.4 | 560 | 13 | AAQ20910 | Thermophilic bacte |
| 32 | 174 | 11.3 | 501 | 16 | AAI00881 | Thermus flavus rec |
| 33 | 174 | 11.3 | 501 | 17 | AAI17719 | Thermus flavus AT |
| 34 | 174 | 11.3 | 552 | 16 | AAI00882 | Thermus thermophil |
| 35 | 174 | 11.3 | 552 | 17 | AAI17720 | Thermus thermophil |
| 36 | 162.5 | 10.5 | 644 | 21 | AAI12943 | Aspergillus oryzae |
| 37 | 160 | 10.3 | 8088 | 21 | AAI245540 | Complete nucleotid |
| 38 | 160 | 10.3 | 8535 | 15 | AAQ73731 | GaLV SEATO genome |
| 39 | 157 | 10.2 | 8060 | 18 | AAI74811 | Porcine retrovirus |
| c 40 | 157 | 10.2 | 8060 | 22 | AAI77725 | Tsukuba-1 cDNA. P |
| 41 | 156 | 10.1 | 8202 | 17 | AAI09280 | Novel AMP/MCF viru |
| 42 | 155 | 10.0 | 8440 | 17 | AAI13265 | Pely F6A provirus |
| 43 | 154 | 10.0 | 3612 | 25 | ABX57820 | MoMuV retrovirus p |
| 44 | 154 | 10.0 | 3643 | 21 | AAI87841 | MLV reverse trans |
| 45 | 154 | 10.0 | 5178 | 21 | AAI15660 | Gibbon leukaemia v |

ALIGNMENTS

RESULT 1
ABV30263
ID ABV30263 standard; cDNA; 2129 BP.
XX
AC ABV30263;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 30254.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX
OS pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200160860-A2.
XX

PD 23-AUG-2001.
 XX 20-FEB-2001; 2001WO-US05171.
 XX 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX Schlegel R, Endege WO, Monahan JE;
 XX WPI; 2001-662795/76.
 XX Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 XX
 XX Claim 1; Page 6567-6568; 11750pp; English.
 XX The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
 XX
 XX Sequence 2129 BP; 609 A; 510 C; 578 G; 427 T; 5 other;
 SQ

Alignment Scores:
 Pred. No.: 5,9e-126 Length: 2129
 Score: 1826.00 Matches: 282
 Percent Similarity: 99.30% Conservative: 2
 Best Local Similarity: 98.60% Mismatches: 2
 Query Match: 98.71% Indels: 0
 DB: 23 Gaps: 0

US-10-054-313-1 (1-286) x ABV30263 (1-2129)
 QY 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArg 20
 DB 112 ATGAGCTGGCTTCTGTTCTCGCCACAGAGTCGCTTGGCCGCTTGGCTGGCCGCGC 171
 QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgGlyThrGlyValPhe 40
 DB 172 GGCCTCTGGCGGGTTCGGGATGTTCTAAGCGGTGAGGAGGGCGGCAAGACCGGGTCTTT 231
 QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
 DB 232 CTGACCTGGAAATGAGTGCAGAGCACAGGTGGACCGGTTTCTGCTGCCAGATTTAAGAAG 291
 QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
 DB 292 TTGGCCACAGAGATGAGGCGCTGGGCTTGTTCAGGAAATCTGCAAGCCGCGGAAGTTCA 351
 QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
 DB 352 GAAGGCGATGAATCAACATGCAGACAAAGATCGAGCGCGCAAGCCAGCAGCGACTCCGT 411
 QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120

Db 412 GAGCCACTGGATGGAGATGGACATGAAGACCGCAGAGCCGTATGCAAGCACATGAAGCGG 471
 QY 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
 Db 472 AGCGTGGAGCGCGCGCTCCAGTTAGCAGAGACAGTGTTCCTACATGGAGACTTCGTC 531
 QY 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160
 Db 532 GTGCTCTACACTGATGGCTGCTGCTCCAGTATGGCGTAGAAGCGCGCAGCAGGAATC 591
 QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLysArgLeuProGlyArgGln 180
 Db 592 GCGGTTTACTGGGGCGCAGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGCGGCAG 651
 QY 181 ThrAsnGlnArgAlaGluLeuHisAlaAlaCysLysAlaAlaGluGlnAlaLysThrGln 200
 Db 652 ACAACCAAGAGCGGAAATTCATGCAGCTCGAAAGCCATTGAACAAGCAAGACTCAA 711
 QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
 Db 712 AACATCAATAAAGCTGCTCTGTATACAGACAGTATGTTTACGATAATGGTATAACTAAC 771
 QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
 Db 772 TGGTTCAGAGTTGGAAGAAAATGGGTGGAAGACAAAGTGCAGGGAAGAGGTGATCAAC 831
 QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
 Db 832 AAAGAGGACTTCTGGCACTGGAGAGGCTTACCAGGGGATGGACATTGATGGATGCAT 891
 QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
 Db 892 GTTCCTGTCATTCGGGATTTATAGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA 951
 QY 281 AlalysGlnSerGluAsp 286
 Db 952 GCTAACCAATCGAAGAC 969
 RESULT 2
 AAZ51265
 ID AAZ51265 standard; cDNA; 1150 BP.
 XX AC AAZ51265;
 XX DT 06-JUN-2000 (first entry)
 XX DE Human RNA-associated protein-16 (RNAAP-16) encoding cDNA.
 XX KW RNA-associated protein; RNAAP; human; clone 2073417; cytostatic;
 KW immunosuppressive; antiinflammatory; keratolytic; neuroprotective;
 KW antiarteriosclerotic; hepatotropic; antipsoriatic; virucide; anti-HIV;
 KW antiallergic; antineumatic; antiarthritic; ophthalmological; autoimmune;
 KW antimicrobial; cell proliferative disorder; inflammation; cirrhosis;
 KW actinic keratosis; burns; arteriosclerosis; artherosclerosis;
 KW hepatitis; myelofibrosis; primary thrombocythemia; psoriasis; cancer;
 KW mixed connective tissue disease; MCTD; HIV; uveitis; Crohn's disease;
 KW allergy; rheumatoid arthritis; parasitic infection; ss.
 XX OS Homo sapiens.
 XX Key Location/Qualifiers
 PH 85..945
 FT CDS /*tag= a
 FT /product= "Human RNA-associated protein-16"
 FT /note= "Derived from ISLW0701 library"
 FT 85..138
 FT sig_peptide /*tag= b
 FT 139..942
 FT mat_peptide /*tag= c
 FT /product= "Mature human RNA-associated protein-16"
 XX WO200011171-A2.
 PN
 XX

PD 02-MAR-2000.
XX 20-AUG-1999; 99WO-US19361.
XX 21-AUG-1998; 98US-0097550.
PR 12-JAN-1999; 99US-0115639.
XX (INCY-) INCYTE PHARM INC.
XX Hillman JL, Yue H, Tang YT, Corley NC, Guegler KJ, Gorgone GA;
PI Patterson C, Baughn MR, Lal P, Bandman O, Reddy R, Azimzal Y;
PI Shih LL, Yang J, Lu DAM;
XX
DR WPI; 2000-237651/20.
DR P-PSDB; AAY70235.
XX
PT Human RNA-associated proteins useful in diagnosing, treating and
PT preventing cell proliferative, autoimmune, inflammatory and infectious
PT disorders -
XX
PS Claim 9; page 117-118; 123pp; English.
XX
CC The present sequence is the cDNA encoding human RNA-associated protein-16
CC (RNAAP-16), identified in Incyte clone 2073417, derived from ISiNOT01
CC library. It is expressed in nervous, reproductive, gastrointestinal,
CC cardiovascular and haematopoietic/immune tissues. It has cytostatic,
CC immunosuppressive, anti-inflammatory, anti-arteriosclerotic, hepatotropic,
CC keratolytic, neuroprotective, antipsoriatic, anti-HIV, anti-allergic,
CC antirheumatic, virucide, antiarthritic, ophthalmological and antimicrobial
CC activity. RNAAP antibodies are useful for diagnosis of diseases
CC associated with altered expression or activity of RNAAP. It is used to
CC treat cell proliferative, autoimmune, inflammatory and infectious
CC disorders, like actinic keratosis, bursitis, arteriosclerosis,
CC artherosclerosis, cirrhosis, hepatitis, myelofibrosis, mixed connective
CC tissue disease (MCTD), psoriasis, primary thrombocythemia and cancer,
CC HIV, allergies, rheumatoid arthritis, uveitis, Crohn's disease, and
CC bacterial, viral and parasitic infections.
XX
SQ Sequence 1150 BP; 302 A; 250 C; 352 G; 246 T; 0 other;

Alignment Scores:
Pred. No.: 5,1e-126 Length: 1150
Score: 1523.00 Matches: 281
Percent Similarity: 29.30% Conservative: 3
Best Local Similarity: 98.25% Mismatches: 2
Query Match: 98.51% Indels: 0
DB: 21 Gaps: 0

US-10-054-313-1 (1-286) x AA251265 (1-1150)

QY 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
DB 85 ATGAGCTGGCTTCTCTTCTGGCCACAGATGCGCTTGGCCGCTTGGCCGCGC 144
QY 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgArgGlyValGlyValPhe 40
DB 145 GGCCTCGCGGGTTCGGGATGTTCTATCGCTGAGGAGCGCGCAGACCGGGTCTTT 204
QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60
DB 205 CTGACCTGGAATGAGTGCAGACAGACAGTGGACCGCTTCTGCTGCGAGATTAAAGAAG 264
QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
DB 265 TTTCACACAGAGATGAGCGCTGGGCCCTTTGTACAGAAATCTGCAAGCCCGAAGTTTCA 324
QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
DB 325 GAAGGGCATGAAATCAACATGGAACAAGATCGGAGCGGAAAGCCAGCAGCGACTCGGT 384
QY 101 GluProLeuAspGlyHisGlyHisGluSerAlaGlnProTrpAlaLysHisMetLysPro 120
DB 385 GAGCCATGGATGGAGATGGACATGAAGCCGAGCGCGGTATGCAAGCATGAAAGCCG 444

QY 121 SerValGluProAlaProValSerArgAspThrPheSerTrpMetGlyAspPheVal 140
DB 445 AGCATGGAGCGCGGCTCCAGTTAGCAGAGACACGTTTTCTTACATGGGACATTCGTC 504
QY 141 ValValTrpThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160
DB 505 GTGCTCTACACTGATGGTCTCTCCAGTAATGGCGTAGAAGCGCGGAGGAGATC 564
QY 161 GlyValTrpThrGlyProGlyHisProLeuAsnValGlyLysArgLeuProGlyArgGln 180
DB 565 GCGTTTACTGGGGCCAGGCCATCTTTAAATGTAGGCAATTAGACTTCTCTGGGCGGAG 624
QY 181 ThrAsnGlnArgAlaGluLeuHisAlaAlaCysLysAlaLeuLeuGluGlnAlaLysThrGln 200
DB 625 ACAACCAAGAGCGGAAATTCATGAGCTGCAAGCCATTGAACAAGCAAGACTCAA 684
QY 201 AsnIleAsnLysLeuValLeuTrpThrAspSerMetPheThrIleAsnGlyLysThrAsn 220
DB 685 AACATCAATAAAGTCTCTGTATATACAGACAGTATGTTTACGATAAATGGTATACTAAC 744
QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
DB 745 TGGTTTCAAGTTGGAAGAAATGGGTGGAAGACAGTGCAGGGAAGAGGTGATCAAC 804
QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
DB 805 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCAGGGGATGCACATTCACTGGATGCAT 864
QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGluGly 280
DB 865 GTTCTTGGTCACTGGGATTTATAGCAATGAAGAGCTGACAGATTGCCAGAGAAGGA 924
QY 281 AlaLysGlnSerGluAsp 286
DB 925 GCTAAACAATCGAAGAC 942

RESULT 3
ABV45787
ID ABV45787 standard; cDNA; 545 BP.
XX AC ABV45787;
XX AC ABV45787;
DT 16-SEP-2002 (first entry)
XX DE Human prostate expression marker cDNA 45778.
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KW pharmacogenomic marker; gene; ss.
XX OS Homo sapiens.
XX PN WO200160960-A2.
XX XX 23-AUG-2001.
XX PF 20-FEB-2001; 2001WO-US05171.
XX PR 17-FEB-2000; 2000US-183319P.
XX PR 16-MAR-2000; 2000US-189862P.
XX PR 25-MAY-2000; 2000US-207454P.
XX PR 09-JUN-2000; 2000US-211314P.
XX PR 18-JUL-2000; 2000US-219007P.
XX PR 13-DEC-2000; 2000US-255281P.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful

PT for detecting presence of prostate cancer, stage of prostate cancer -

PS Claim 1; Page 9051; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising

CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

CC specification or its complement. (I) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;

CC (c) assessing the efficacy of a test compound to inhibit prostate

CC cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer

CC in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a

CC patient;

CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 545 BP; 136 A; 129 C; 178 G; 102 T; 0 other;

SQ

Alignment Scores:

| | | | |
|------------------------|----------|---------------|-----|
| Pred. No.: | 1,628-71 | Length: | 545 |
| Score: | 905.00 | Matches: | 167 |
| Percent Similarity: | 94.44% | Conservative: | 3 |
| Best Local Similarity: | 92.78% | Mismatches: | 10 |
| Query Match: | 58.54% | Indels: | 0 |
| DB: | 23 | Gaps: | 0 |

US-10-054-313-1 (1-286) x ABV45787 (1-545)

QY 13 LeuAlaLeuProCysArgArgGlySerArgGlyPheGlyMetPheTyrAlaValArg 32

DB 4 ATTGGAGCTCCACCGGG 63

QY 33 ArgGlyArgGlyPheGlyValPheLeuThrTrpAenGluCysArgAlaGlnValAspArg 52

DB 64 AGGGGGCGGAAAGTGGGGTCTTTCTGACCTGGATGAGTGCAGAGCAGAGTGGACCG 123

QY 53 PheProAlaAlaArgPheGlyLysPheAlaThrGluAspGluAlaTrpAlaPheValArg 72

DB 124 TTTCTCTGCTCCAGATTAAAGATTGGCCACAGAGGATGAGGCTGGGCTTTGTCAGG 183

QY 73 LysSerAlaSerProGluValSerGluGlyHisGluAenGlnHisGlyClnGluSerGlu 92

DB 184 AAATCTGCAAGCCCGGAAGTTTCAGAGGGCGCATGAAATCAACATGGACAGATCGGAG 243

QY 93 AlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGlyHisGluSerAlaGln 112

DB 244 GCGAAAGCCAGCAAGCGACTCCGTGAGCCACTGGATGGAGATGGACATGAAGCGCAGAG 303

QY 113 ProTyrAlaLysHisMetLysProSerValGluProAlaProValSerArgAspThr 132

DB 304 CCGTATGCAAGCAGATGAAGCGAGCGGAGCGCGCCCTCCAGTTAGCAGACACAG 363

QY 133 PheSerTyrMetGlyAspPheValValValTyrThrAspGlyCysCysSerSerAenGly 152

DB 364 TTTTCTACATGGAGACTTCGTCGTCTACACTGATGGCTGCTCCAGTAATGGG 423

QY 153 ArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGlyHisProLeuAenVal 172

DB 424 CGTAGAAGCGCGGAGCAGGAATCGCGTTTACTGGGGGGCGGGCCATCTTTAAATGTA 483

QY 173 GlyLysArgLeuProGlyArgGlnThrAenGlnArgAlaGluIleHisAlaLysLys 192

DB 484 GGCATTAGACTTCTGGGGGGGAGACAAACCAAGAGCGGGAATTCATGCACCTGCANA 543

RESULT 4

ID ABV15988

XX ABV15988 standard; cDNA; 691 BP.

AC ABV15988;

XX 13-SEP-2002 (first entry)

XX Human prostate expression marker cDNA 15979.

DB

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX pharmacogenomic marker; Gene; ss.

XX Homo sapiens.

XX WO200160860-A2.

XX 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US05171.

XX 17-FEB-2000; 2000US-183319P.

XX 16-MAR-2000; 2000US-19862P.

XX 25-MAY-2000; 2000US-207454P.

XX 09-JUN-2000; 2000US-211314P.

XX 18-JUL-2000; 2000US-219007P.

XX 13-DEC-2000; 2000US-255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JE;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of

PT prostate cells and correlating with presence of prostate cancer, useful

PT for detecting presence of prostate cancer, stage of prostate cancer -

XX Claim 1; Page 2678; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising

CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

CC specification or its complement. (I) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;

CC (c) assessing the efficacy of a test compound to inhibit prostate

CC cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer

CC in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a

CC patient;

CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 691 BP; 186 A; 155 C; 199 G; 144 T; 7 other;

SQ

Alignment Scores:

| | | | |
|------------------------|----------|---------------|-----|
| Pred. No.: | 3,36e-50 | Length: | 691 |
| Score: | 666.50 | Matches: | 159 |
| Percent Similarity: | 73.82% | Conservative: | 13 |
| Best Local Similarity: | 68.24% | Mismatches: | 34 |
| Query Match: | 43.11% | Indels: | 28 |
| DB: | 23 | Gaps: | 8 |

US-10-054-313-1 (1-286) x ABV15988 (1-691)

QY 24 GlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPheLeuThrTrp 43

DB 3 GGGTTCCGGATGTTCTATCGGTGAGGAGGGCCGCAAGACTGGGGTCTTCTGACCTGG 62

QY 44 AsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysPheAlaThr 63

DB 63 AATGAGTGCAGACACAGAGTGGACCGGTTTCTCTGCTCCAGATTAAAGAGTTTGCCACA 122

QY 64 GluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHis 83

Db 123 GAGGATGAGGCTGGGCTTTGTTCAGGAATCTCGAAGCCGGAAGTTTCAGAGGGCAT 182
 Qy 84 GluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeu 103
 Db 183 GAAATCAACATGACAGAAATCGAGCGGAAACCCAGACGACTCCGTGAGCCACTG 242
 Qy 104 AspGlyAspGlyHisGluSerAlaGln-ProTyrAlaLysHis-MetLysProSerVal- 122
 Db 243 GATGAGATGACATGAAGCGCAGAGCCCGTATGCAAGCACCATGAAGCCGAGCGGT 302
 Qy 123 --GluProAlaProProValSerArgAspThr-PheSerTyrMet-GlyAspPheVal-V 141
 Db 303 GGAGCCCGCGCTCCAGTTAGCAGACACCGTTTCTACATGGGAGACTTCGTTCG 362
 Qy 141 alValTyrThrAspGly-CysCysSerSerAsnGlyArgArgLys-----ProArgAla 158
 Db 363 TCGTCTACATGATGGCTTGTCTCCAGTAAATGGGCGGTANGAAGCCCGCCGAGGCAG 422
 Qy 159 GlyIleGlyValTyrTrp---GlyProGlyHisProLeuAsn---ValGlyIleArgLeu 176
 Db 423 GGAATCGGCGTTTACTGGGGGGGCGCGGCCCATCTTTAAATGTTAGGCCATTAAAGACTT 482
 Qy 177 Pro-----GlyArg-GlnThrAsnGlnArgAlaGluIleHisAlaLa-----CysLY 192
 Db 483 CCTTGGCGCGCAGGACACAAACCAAGAGAAAGCGGNAATTTATGGCCANGCCTTGCAA 542
 Qy 192 sAlaIleGluGlnAlaLysThrGlnAsnIleAsnLys----- 204
 Db 543 AAGCCATTGAAACCAAGCGCAAGAGACTTCATAAAACCATTCATAATAAAACCTGGG 602
 Qy 205 -LeuValLeuTyrThrAspSerMetPheThrIle-----AsnGlyIleThrAsnTr 221
 Db 603 TTTCTGTGTTATCCAGACAA-GTTTNTTGTTTTACCGAATTAATAATGGTNTTTAAACTT 661
 Qy 221 pValGlnGlyTrpLysLysAsnGly 229
 Db 662 AACCTGGGTTTCAANGGTTNGG 686
 RESULT 5
 ID AAH30241 standard; cdna; 389 BP.
 AC AAH30241;
 XX 27-JUL-2001 (first entry)
 DT Human colon cancer cell line Km12L4-A cDNA library derived sequence #175.
 DE Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;
 KW detection; colon cancer cell line Km12L4-A; ss.
 KW Homo sapiens.
 OS
 XX WO200018916-A2.
 PN 06-APR-2000.
 XX 23-SEP-1999; 99WO-US22226.
 PF 28-SEP-1998; 98US-0102161.
 PR 28-SEP-1998; 98US-0102180.
 PR 29-SEP-1998; 98US-0102380.
 PR 08-OCT-1998; 98US-0103815.
 PR 27-OCT-1998; 98US-0105877.
 XX (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX Williams LT, Escobedo J, Innis WA, Garcia PD, Sudduth-Klinger J;
 PI Reinhard K, Glese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
 PI Lemson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
 PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
 XX

DR WPI; 2000-293155/25.
 XX Polynucleotide library comprising 1079 defined sequences, useful in
 PT the form of an array to detect cancer or susceptibility to cancer -
 XX Claim 1; Page 233; 502pp; English.
 XX The present invention describes a library of polynucleotides comprising
 CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
 CC are: (1) an isolated polynucleotide (I) having at least 90% identity to
 CC one of the 1079 sequences; (2) a recombinant host cell containing (1);
 CC (3) an isolated polypeptide (II) encoded by (1); (4) an antibody that
 CC specifically binds to (II); (5) a vector comprising (1); and (6) a method
 CC of detecting differentially expressed genes correlated with a cancerous
 CC state of a mammalian cell comprising detecting a gene product encoded by
 CC 65 of the 1079 sequences given in the specification. The polynucleotides
 CC are used to monitor patients having (or susceptible) to cancer to detect
 CC potentially malignant events at a molecular level before they are
 CC detectable at a gross morphological level. The polynucleotides are also
 CC useful for monitoring the efficacy of various therapies and preventive
 CC interventions. Polynucleotide probes based on the disclosed sequences
 CC are useful for chromosome mapping and detection of transcription levels.
 CC The 1079 polynucleotide sequences were derived from a human colon cancer
 CC cell line Km12L4-A cDNA library.
 XX Sequence 389 BP; 78 A; 101 C; 130 G; 80 T; 0 other;
 SQ

Alignment Scores: Length: 389
 Pred. No.: 1.8e-36 Matches: 104
 Score: 508.50
 Percent Similarity: 56.15% Conservative: 1
 Best Local Similarity: 55.61% Mismatches: 0
 Query Match: 32.89% Indels: 82
 DB: 21 Gaps: 1

US-10-054-313-1 (1-286) x AAH30241 (1-389)

Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArg 20
 Db 74 ATGAGCTGGCTTCTGTCTTGGCCACAGATCGCTTGGCCCTTCCCTCCGCGCGC 133
 Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
 Db 134 GGCTCTCGGGGTTCGGATGTTCTATGCGTGAGGAGGCGCGCAAGACCGGGTCTTT 193
 Qy 41 LeuThrTrpAsnGlnCysArgAlaGlnValAspArgPheProAlaAlaArg-PheLysLys 60
 Db 194 CTGACCTGGAATGAGTGCAGACAGACAGGTGGACCGGTTTCTCTGTCAGACTTC----- 248
 Qy 60 sPheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSe 80
 Db 248 ----- 248
 Qy 80 rGluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuAr 100
 Db 248 ----- 248
 Qy 100 gGluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPr 120
 Db 248 ----- 248
 Qy 120 oSerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVa 140
 Db 249 -----GT 250
 Qy 140 lValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
 Db 251 CGTCTGTACACTGATGGCTGCTCTCCAGTAAATGGCGTAGAAGCCCGCAGCAGGAAT 310
 Qy 160 eGlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgG 180
 Db 311 CGGGCTTTACTGGGGCGCGGCCCATCTCTTTAAATGTAGGCATTAGACTTCTCTGGCGGCA 370

QY 180 nThrAsnGlnArgAlaGlu 186
 DB 371 GACAAACCAAGAGCGGAA 389

RESULT 6
 ABX41815
 ID ABX41815 standard; cDNA; 473 BP.
 XX
 AC ABX41815;
 XX
 DT 20-FEB-2003 (first entry)
 XX
 DE Bovine EST associated with lactation/muscle/fat deposition #6980.
 XX
 XX Bovine, ss; EST; expressed sequence tag; lactation; LMFD;
 KW muscle deposition; fat deposition; genome mapping; gene identification;
 KW gene analysis; cattle breeding.
 XX
 OS Bos Taurus.
 XX
 PN US2002137139-A1.
 XX
 PD 26-SEP-2002.
 XX
 PF 24-SEP-2001; 2001US-0960352.
 XX
 PR 12-JAN-1999; 99US-115707P.
 PR 11-JAN-2000; 2000US-0460902.
 XX
 XX (BYAT/) BYATT J C.
 PA (NATH/) MATHIALAGAN N.
 PA (TAON/) TAO N.
 PA (WARR/) WARREN W C.
 XX
 PI Byatt JC, Mathialagan N, Tao N, Warren WC;
 XX
 DR WPI; 2003-110599/10.
 XX
 XX New nucleic acid associated with lactation, and muscle and fat
 PT deposition, useful for genome mapping, gene identification and
 PT analysis, cattle breeding, or for genetically improving cattle
 XX
 PS Claim 2; SEQ ID NO 6980; 245pp; English.
 XX
 CC The invention relates to a purified nucleic acid molecule associated with
 CC lactation or muscle and fat deposition (designated LMFD), derived
 CC from cattle, and the LMFD nucleic acid can specifically hybridise to a
 CC second nucleic acid molecule comprising any of 15112 nucleotide
 CC sequences, appearing as ABX34836-ABX49947, or complements of them.
 CC Also included are: (1) a transformed cell having a nucleic acid
 CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-
 CC translated sequence that functions in the cell to cause termination of
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end
 CC of the mRNA molecule; and (2) determining a level or pattern of a
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
 CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its
 CC complement or fragment) with a complementary nucleic acid molecule
 CC obtained from the bovine cell or tissue, where hybridisation between the
 CC marker nucleic acid and the complementary nucleic acid permits the
 CC detection of the molecule; and (b) detecting the level or pattern of the
 CC complementary nucleic acid, where the detection of the complementary
 CC nucleic acid is predictive of the level or pattern of the molecule.
 CC The LMFD nucleic acid is used for determining a level or pattern
 CC of a molecule in a bovine cell or tissue. It is useful for genome
 CC mapping, gene identification and analysis, cattle breeding, preparation
 CC of constructs for use in cattle gene expression or for genetically
 CC improving cattle. The present sequence is one of the 15112 bovine
 CC LMFD EST (expressed sequence tag) nucleic acids.
 CC Note: The present sequence was not shown in the specification but
 CC was obtained in electronic format from the USPTO web site:
 CC seqdata.uspto.gov/sequence.html?DocID=20020137139.
 XX
 SQ Sequence 473 BP; 129 A; 109 C; 140 G; 94 T; 1 other;

Alignment Scores: 9.51e-30 Length: 473
 Pred. No.: 434.00 Matches: 79
 Score: 88.78% Conservative: 8
 Percent Similarity: 80.61% Mismatches: 11
 Best Local Similarity: 28.07% Indels: 0
 Query Match: 25 Gaps: 0
 DB: 25

US-10-054-313-1 (1-286) x ABX41815 (1-473)

QY 189 AlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyr 208
 DB 79 GCAGCCTGCAAGCCATAGAGCAAGCTAAGGCTCAGGACATCACTAAGCTGGTTCCTCTAC 138
 QY 209 ThrAspSerMetPheThrIleAsnGlyIleThrAsnTyrValGlnGlyTyrLysAsn 228
 DB 139 ACAGACAGCATGTTTACCATCAATGGCATCACCACTGGTGAAGGCTGGAGCAGAAAT 198
 QY 229 GlyTyrLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeuGlu 248
 DB 199 GGCTGGAGGACAGCAGCCGGGAAGAGGTGACCAACAAGGAGGACTTCGCGGAGCTGGAG 258
 QY 249 ArgLeuThrGlnGlyMetAspIleGlnTyrMetHisValProGlyHisSerGlyPheIle 268
 DB 259 CGGCTGGCGGGGCGCATGGACATTCAGTGGATGCAGCTTCCTGCCATTCGGGGTTTAAA 318
 QY 269 GlyAsnGluGlnAlaAspArgLeuAlaArgGluGlyValAlaLysGlnSerGluAsp 286
 DB 319 GGCAATGAAAGGCGGACAGACTATCGAGAGAGGCTCGAAACACTCCGCAGAC 372

RESULT 7
 ABL03247
 ID ABL03247 standard; cDNA; 1071 BP.
 XX
 AC ABL03247;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 4223.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEXE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-658860/75.
 DR P-PSDB; ABB59144.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 4223; 21pp + Sequence Listing; English.
 CC
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB70272).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1071 BP; 320 A; 229 C; 292 G; 230 T; 0 other;

Alignment Scores:

Pred. No.: 3,036-28 Length: 1071
 Score: 422.00 Matches: 107
 Percent Similarity: 44.37% Conservative: 31
 Best Local Similarity: 34.41% Mismatches: 113
 Query Match: 27.30% Indels: 60
 DB: 23 Gaps: 7

US-10-054-313-1 (1-286) x ABL03247 (1-1071)

QY 28 PheTyrAlaValArgArgGlyArgLysThrGlyValPheLeuThrTrpAsnGluCysArg 47
 DB 52 TTTTACGCTGTAGCCAGTGGCGGGCTCTGGAGCTACGGCTCTTGGCGAGATGCGAG 111
 QY 48 AlaGlnValAspArgPheProAlaAlaArgPheLysLysPheAlaThrGluAspGluAla 67
 DB 112 GAGCAGGTCAAGGGTTTCAAGAACGCCAAATACAAAGAGTTTAAACACACGCCAGGAGCG 171
 QY 68 TrpAlaPheValArg-----LysSerAlaSerProGlu----- 78
 DB 172 GATCAGTTTCGTTAATGGTGCAGTGTATGTCCTCCGAGATGTGGCGTCCCGTGGGC 231
 QY 78 ----- 78
 DB 232 AAGGAAAGCGCTCACTGCCAGTTGGAAAACAGTATCGAAGTGAATAAGAACCCCAAG 291
 QY 79 ---ValSerGluGlyHisGluAsnGlnHisGly----- 88
 DB 292 TACACAGATGATGGCGCCGAGAGATCAGCAGCTAGCCGAGGATGATCTGAATGCTGCC 351
 QY 89 -----GlnGluSerGluAlaLysPro----- 95
 DB 352 ATGACGAGGTTGAGGAGATCCAAAGCCATCTAACAGTAGCAATTTACCAGACATTC 411
 QY 96 GlyLysArgLeuArgGluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAla 115
 DB 412 AATCGCAAGCGGAGCGCACACACCGCGCGATAGCGAAACAGATCCACGTCATGCC 471
 QY 116 LysHisMetLysProSerValGluProAlaProValSerArgAspThrPheSerTyr 135
 DB 472 TCTCAGTCTCA-----GAAGCCACGAGACTCAAGCAAGTGGTGCCTTCAGTTT 522
 QY 136 MetGlyAsp-----PheValValValTyrThrAspGlyCysCysSerSerAsnGly 152
 DB 523 GAATCGATGACGAGCGGCTATCTATTGTGTACACAGATGGCTCTCGTCAAGCAACGGA 582
 QY 153 ArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGlyHisProLeuAsnVal 172
 DB 583 CGAGCGGAGCGCTGTGCCGGCTATGCGGTTATTTCGGCAAGATCAACAGCTAAACGCA 642
 QY 173 GlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLys 192
 DB 643 GCCAAGCCGCGAAGGAGCGCTTACAAATATGTAGTGAATACAAAGCGGCATTCAT 702
 QY 193 AlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMet 212
 DB 703 GCATTAAACAGCTCTTGACTTGGGAATACAGAAGCTGTGCATCAGCACAGACTCTCAG 762
 QY 213 PheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThr 232
 DB 763 TTTTGTATCACTCCATACAGCTGTGGTTCGGGTGGAAAAAAGGATGGAGCTA 822
 QY 233 SerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGln 252

Db 823 AAGAACAAATCAGCCTGTTAAACACGTTGTGACTTCAAGGAACTGGATAAAGTCTCAG 882
 QY 253 -----GlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsn 270
 DB 883 GAGAACAAATCATCACCTGTAATGGAATGAACTATGTGGAGGCCACAAAGGCGCATAGAGGCAAT 942
 QY 271 GluGluAlaAspArgLeuAlaArgGluGlyAla 281
 DB 943 GAAATGCGACAAATTTGGCGCGCAAGGATCC 975

RESULT 8 ABX38735

ID ABX38735 standard; cDNA; 407 BP.

AC ABX38735;

XX 20-FEB-2003 (first entry)

DE Bovine EST associated with lactation/muscle/fat deposition #3900.

KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
 KW muscle deposition; fat deposition; genome mapping; gene identification;
 KW gene analysis; cattle breeding.

OS Bos Taurus.

XX US2002137139-A1.

XX 26-SEP-2002.

XX 24-SEP-2001; 2001US-0960352.

XX 12-JAN-1999; 99US-115707P.

PR 11-JAN-2000; 2000US-0480902.

XX (BYAT/) BYATT J C.

PA (MATH/) MATHIALAGAN N.

PA (TAON/) TAO N.

PA (WARR/) WARREN W C.

XX Byatt JC, Mathialagan N, Tao N, Warren WC;

XX WPT; 2003-110599/10.

XX New nucleic acid associated with lactation, and muscle and fat
 deposition, useful for genome mapping, gene identification and
 analysis, cattle breeding, or for genetically improving cattle

PS Claim 2; SEQ ID No 3900; 245pp; English.

XX The invention relates to a purified nucleic acid molecule associated with
 CC lactation or muscle and fat deposition (designated LMFD), derived
 CC from cattle, and the LMFD nucleic acid can specifically hybridise to a
 CC second nucleic acid molecule comprising any of 15112 nucleotide
 CC sequences, appearing as ABX34836-ABX49947, or complements of them.
 CC Also included are: (1) a transformed cell having a nucleic acid
 CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-
 CC translated sequence that functions in the cell to cause termination of
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end
 CC of the mRNA molecule; and (2) determining a level or pattern of a
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
 CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its
 CC complement or fragment) with a complementary nucleic acid molecule
 CC obtained from the bovine cell or tissue, where hybridisation between the
 CC marker nucleic acid and the complementary nucleic acid permits the
 CC detection of the molecule; and (b) detecting the level or pattern of the
 CC complementary nucleic acid, where the detection of the complementary
 CC nucleic acid is predictive of the level or pattern of the molecule.
 CC The LMFD nucleic acid is used for determining a level or pattern
 CC of a molecule in a bovine cell or tissue. It is useful for genome
 CC mapping, gene identification and analysis, cattle breeding, preparation
 CC of constructs for use in cattle gene expression, or for genetically
 CC improving cattle. The present sequence is one of the 15112 bovine

QY 175 gLeuProGlyArgGlnThrAsn-----GlnArgala 185


```

Db      2170 AGACAAATTGGCGGCGCAAGGATCC 2194
RESULT 11
AAF28553/c
ID      AAF28553 standard; DNA; 119211 BP.
XX
AC      AAF28553;
XX
DT      04-APR-2001 (first entry)
XX
DE      Genomic fragment #40.
XX
KW      Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW      bronchopulmonary; endocarditis; meningitis; ss.
XX
OS      Moraxella catarrhalis.
XX
PN      WO200078968-A2.
XX
PD      28-DEC-2000.
XX
PF      16-JUN-2000; 2000WO-US16649.
XX
PR      18-JUN-1999; 99US-0140121.
XX
PA      (INCY-) INCYTE GENOMICS INC.
XX
PI      Lagace RE, Patterson C, Berg KL;
XX
PS      WPI; 2001-041427/05.
XX
PT      Genomic library for identifying diagnostic and therapeutic
PT      compositions, and for identifying virulence factors, regulatory
PT      elements and drug targets, comprises Moraxella catarrhalis nucleic
PT      acids -
XX
PS      Claim 1; Page 459-485; 545pp; English.
XX
CC      The present invention relates to a Moraxella catarrhalis genomic library
CC      comprising of a combination of 41 nucleic acid molecules (see
CC      AAF28514-AAF28554). The library has a number of uses described in the
CC      specification e.g. is useful for identifying diagnostic and therapeutic
CC      compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC      aerobic gram-negative diplococcus, normally found among the bacterial
CC      flora of human upper airways. M. catarrhalis is known to cause acute,
CC      localised infections such as otitis media, sinusitis and bronchopulmonary
CC      infection and life-threatening, systemic diseases including endocarditis
CC      and meningitis.
XX
SQ      Sequence 119211 BP; 34729 A; 23773 C; 25664 G; 35041 T; 4 other;

Alignment Scores:
Pred. No.:      2.15e-12      Length:      119211
Score:          272.00      Matches:      62
Percent Similarity: 52.23%      Conservative: 20
Best Local Similarity: 39.49%      Mismatches: 51
Query Match:    17.59%      Indels:      24
DB:             22          Gaps:         5

US-10-054-313-1 (1-286) x AAF28553 (1-119211)
QY      136 MetGlyAspPheValValValValThrAspGlyCysSerSerAsnGlyArgArgLys 155
Db      45182 ATGAGCCAACTTAATATCGCTTATACGCGCGCCCTGTAAAGGCAATGAAAAACAAGGC 45123
QY      156 ProArgla---GlyIleGlyValtyr----- 163
Db      45122 GTATCTCAGCGCGTTGGCGCGTGTATTTCATTATTTCATGCGATGACCGGATCTCG 45063
QY      164 TrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrAsnGln 183
Db      45062 TGGGGC---GGTGAGCCT-----GATACGACCAATAAT 45033

```

```

QY      184 ArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsn 203
Db      45032 CGCATGGAATTGATGCTGCCATCAGCTTTGGAGGCAACGCTGCACAG-----ATT 44979
QY      204 LysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGln 223
Db      44978 CCTCTGCAACTTTGGACAGATTTCAGGCTATGTTAAAGATGGCATACTCAGTGATGGC 44919
QY      224 GlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAsp 243
Db      44918 GGTGGAAAGTTGCTGTTGGAAAAAGCAGATGGCAAGCTGCTTAATCAAGACCTA 44859
QY      244 PheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGly 263
Db      44858 TGGCAACGATTGGATCAGCTGACCCAAATCGCATCATGATTGGCAATGATGATCAAGGC 44799
QY      264 HisSerGlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGluGly 280
Db      44798 CATGCTGCTCATGCTGGCAATGAGATGGCAACCACTGGCCCAATAAAGGC 44748

RESULT 12
AAT24248
ID      AAT24248 standard; cDNA to mRNA; 378 BP.
XX
AC      AAT24248;
XX
DT      16-OCT-1996 (first entry)
XX
DE      Human gene signature HUMGS06264.
XX
KW      Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW      human; cloning; mapping; non-biased library; diagnosis; detection;
KW      cell typing; abnormal cell function; ss.
XX
OS      Homo sapiens.
XX
PN      WO9514772-A1.
XX
PD      01-JUN-1995.
XX
PF      11-NOV-1994; 94WO-JP01916.
XX
PR      12-NOV-1993; 93JP-0355504.
XX
PA      (MATS/) MATSUBARA K.
PA      (OKUB/) OKUBO K.
PI      Matsubara K, Okubo K;
XX
PS      WPI; 1995-206931/27.
XX
PT      Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT      for diagnosis of abnormal cell function, by preparing cDNA that
PT      reflects relative abundance of corresp. mRNA in specific human
PT      tissues
XX
PS      Claim 1; Page 1565; 2245pp; Japanese.
XX
CC      A single-stranded DNA (or its complementary strand or the corresp.
CC      double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC      given in AAT19001-T26837 and which is able to hybridise to part of
CC      human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC      sequences were obtained from 3'-directed cDNA libraries prepared
CC      from various human tissues; synthesis of cDNA was initiated from the
CC      3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC      untranslated sequence is unique to a particular mRNA species, almost
CC      all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC      is constructed so as to reflect accurately the relative abundance of
CC      different mRNAs in the particular tissue from which it was derived.
CC      The appearance frequency of a given GS in a cDNA library can be
CC      determined (esp. using primers and probes derived from the GS
CC      sequences) as a means of diagnosing abnormal cell function or for
CC      recognising different cell types.

```

XX SQ Sequence 378 BP; 98 A; 61 C; 106 G; 94 T; 19 other;
 Alignment Scores:
 Pred. No.: 5,46e-13 Length: 378
 Score: 244.00 Matches: 47
 Percent Similarity: 97.92% Conservative: 0
 Best Local Similarity: 97.92% Mismatches: 1
 Query Match: 15.78% Indels: 0
 DB: 16 Gaps: 0

US-10-054-313-1 (1-286) x AAT24248 (1-378)
 QY 239 IleAsnLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrp 258
 DB 2 ATCAACAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCGGGATGACATTCAGTGG 61
 QY 259 MetHisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArg 278
 DB 62 ATGCATGTCNTGTCATTCGGGATTATAGCANTGAAGAAGCTGACAGATTAGCCAGA 121
 QY 279 GluGlyAlaLysGlnSerGluAsp 286
 DB 122 GAAGGAGCTAAACAATCGAGAC 145

RESULT 13
 AAX20568/c
 ID AAX20568 standard; DNA; 7874 BP.
 XX AC AAX20568;
 DT 05-MAY-1999 (first entry)
 XX DE Polynucleotide sequence from the genome of Treponema pallidum.
 XX KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
 KW enzyme production; ds.
 XX OS Treponema pallidum.
 XX PN WO9859034-A2.
 XX PD 30-DEC-1998.
 XX PF 23-JUN-1998; 98WO-US13041.
 XX PR 24-JUN-1997; 97US-0050667.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Fraser CM;
 XX WPI; 1999-081273/07.
 XX PT New isolated Treponema pallidum nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of T. pallidum infections, particularly syphilis
 XX PS Claim 1; Page 560-565; 1150pp; English.
 XX CC AAX20500-21243 represent polynucleotide sequences from the genome of
 CC Treponema pallidum. The sequences can be used for detection,
 CC diagnosis, characterisation, prevention and therapy for T. pallidum
 CC infections, particularly syphilis. They can also be used for detecting
 CC diseases related to Borrelia infections in animals, and for the
 CC production of biosynthetic products such as enzymes.
 XX SQ Sequence 7874 BP; 1590 A; 1949 C; 2387 G; 1929 T; 19 other;
 Alignment Scores:
 Pred. No.: 6.43e-11 Length: 7874
 Score: 239.00 Matches: 52
 Percent Similarity: 46.67% Conservative: 25

Best Local Similarity: 31.52% Mismatches: 58
 Query Match: 15.46% Indels: 30
 DB: 20 Gaps: 2

US-10-054-313-1 (1-286) x AAX20568 (1-7874)
 QY 140 ValValValValThrAspGlyCysCysSerSerAsn- 151
 DB 5784 CTTACCTGTACACTGACGGCCCTGTCTGGGAACCCCGGGTCCGGGGGGGGT 5725
 QY 152 -----GlyArgArgLysProArgAlaGly 159
 DB 5724 GCTCTGTGTCCTCAGACGTACCTTTCTAGAAACCGGACAAACCCGACAGAGTGC 5665
 QY 160 IleGlyValValTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArg 179
 DB 5664 GCATTCACGTAGAACGGCGCAGCGCTACCCG----- 5635
 QY 180 GlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThr 199
 DB 5634 AGCACAACAATCGCATGGAACCTGTGCGCGTCTATCAACGCACCTGCAAGAGGCACACGA 5575
 QY 200 GlnAsnIleAsnLysLeuValLeuValLeuValLeuValLeuValLeuValLeuValLeu 219
 DB 5574 CGCGCTGCAGAGCGGTGTGTGTCACCGACTCCCAATACGTACGCAAAAGGCATCAC 5515
 QY 220 AsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIle 239
 DB 5514 CAGTGGATCCATACCTGGAAACACACACACGCTGGAAACGCTGCCAAACACCGGTCAAG 5455
 QY 240 AsnLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMet 259
 DB 5454 AACAAAGGATTGTGGGAAGCGCTCAGCGCACCTCGCGACGCCCTATCCGTCGAGTGGCG 5395
 QY 260 HisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGlu 279
 DB 5394 TGGTAAAGACATGACGAGACCCGTCACACAACTGTGCACCGGCTAGCCACCGAC 5335
 QY 280 GlyAlaLysGlnSer 284
 DB 5334 GCTGCGCGCGCGCT 5320

RESULT 14
 AAF94380
 ID AAF94380 standard; DNA; 462 BP.
 XX AC AAF94380;
 XX AC AAF94380;
 XX DT 04-JUN-2001 (first entry)
 XX DE Haemophilus influenzae essential bacterial gene SEQ ID NO:71.
 XX KW Haemophilus influenzae; essential bacterial gene; identification;
 KW otitis media; meningitis; upper respiratory tract infection;
 XX infection; antimicrobial; ds.
 OS Haemophilus influenzae.
 XX WO200111033-A2.
 XX PD 15-FEB-2001.
 XX PF 03-AUG-2000; 2000WO-US21176.
 XX PR 04-AUG-1999; 99US-0368382.
 XX PA (ABBO) ABBOTT LAB.
 XX PI Chovan LE, Hessler PE, Reich KA;
 XX WPI; 2001-147511/15.
 XX P-PSDB; AAB88527.

PT Essential bacterial genes from Haemophilus influenzae and methods for
 PT identifying 'essential' genes that may be potential therapeutic targets
 PT

XX Claim 2; Page 116-117; 185pp; English.

XX AAF94345 to AAF94409 represent essential bacterial genes from
 CC Haemophilus influenzae, which encode the proteins given in AAB98492 to
 CC AAB98556. The present invention also describes methods for identifying
 CC essential bacterial genes (i.e. those essential to the survival of a
 CC bacterium) using a transposition system. The methods are used to
 CC identify essential genes from bacteria, especially H. influenzae (which
 CC causes otitis media, meningitis and upper respiratory tract infections)
 CC which may be used as targets for potential antimicrobial agents.
 CC AAF94410 to AAF94416 represent PCR primers used in the exemplification
 CC of the present invention.

XX Sequence 462 BP; 180 A; 74 C; 96 G; 111 T; 1 other;

Alignment Scores:
 Pred. No.: 3,25e-12 Length: 462
 Score: 236.50 Matches: 54
 Percent Similarity: 49.68% Conservative: 23
 Best Local Similarity: 34.84% Mismatches: 49
 Query Match: 15.30% Indels: 29
 DB: 22 Gaps: 4

US-10-054-313-1 (1-286) x AAF94380 (1-462)

QY 140 ValValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgGlyProArgAlaGly 159
 DB 13 ATTGAATTTTACTGATGATCTTCTTAGGTAAT-----CCAGGGCGGGC 60
 QY 160 IleGlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArg 179
 DB 61 -----GGAATTGGTGCCGTATTGCTTATAAA 87
 QY 180 Gln-----ThrAsnGlnArgAlaGluIle 187
 DB 88 CAACATGAAAAAACAACCTCTCAAGGCTATTTCACCAACCCATATATCGAATGGAATTA 147
 QY 188 HisAlaAlaCysLysAlaIleGluAlaLysThrGlnAsnIleAsnLysLeuValLeu 207
 DB 148 CGCGCTGCTCATTTGAAGCAAGCTCTGAAAGCCTGTAAACCAAGATTTATGGATAGCCTTA 321
 QY 248 GluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe 267
 DB 322 GATGAATCCATCCACGCTCTATAAATTAATGGCAATGGTAAAGCCATGCTGACAC 381
 QY 268 IleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAlaLys 282
 DB 382 AGAGAAATGAAATTTGCGATGAATTAGCAAAAAAGGGGCGAGAA 426

RESULT 15

ID ABK64944 standard; DNA; 462 BP.

XX ABK64944;

XX AC

XX 18-JUN-2002 (first entry)

XX DNA encoding Haemophilus influenzae essential gene #36.

XX Essential bacterial gene; antifungal agent; antibacterial agent;

XX antiparasitic agent; insecticidal agent; microbial infection;

XX

KW mucous membrane infection; otitis media; sinusitis; bronchitis;
 KW alveolitis; conjunctivitis; pneumonia; meningitis; epiglottitis;
 KW cellulitis; septic arthritis; gene; ds.

OS Haemophilus influenzae.

PN WO200218601-A2.

XX 07-MAR-2002.

XX 22-AUG-2001; 2001WO-US26245.

XX 25-AUG-2000; 2000US-0649145.

XX (ABBO) ABBOTT LAB.

XX Chovan LE, Hessler PE, Reich KA;

XX WPI; 2002-304258/34.

XX P-PSDB; AAU91458.

XX Essential bacterial genes in Haemophilus influenzae necessary for
 PT bacterium's growth and survival, useful for screening inhibitors of
 PT polypeptides and developing therapeutic agents e.g. antimicrobial
 XX Claim 1; Page 116-117; 185pp; English.

XX The invention describes an essential bacterial gene (I) comprising a
 CC purified polynucleotide isolated from Haemophilus influenzae where (I)
 CC is essential to H. influenzae survival. The encoded polypeptide (II) is
 CC useful for screening substances that function to inhibit essential H.
 CC influenzae polypeptides by contacting (II) with the desired substances
 CC and measuring the response by a screen from specific, enzyme, general,
 CC affinity, phenotypic and binding screen. (I) and (II) are useful in
 CC developing therapeutic agents such as antifungal, antibacterial and
 CC antiparasitic agent, insecticidal agent, and preventive antimicrobial
 CC agents which are effective in preventing microbial infection or useful
 CC in treatment of that particular infection. (I) and (II) may also be
 CC useful in treatment of mucous membrane infections such as otitis media,
 CC sinusitis, bronchitis, alveolitis, conjunctivitis, pneumonia, meningitis,
 CC epiglottitis, cellulitis and septic arthritis. This sequence encodes an
 CC essential H. influenzae gene, described in the invention.

XX SQ Sequence 462 BP; 180 A; 74 C; 96 G; 111 T; 1 other;

Alignment Scores:

Pred. No.: 3,25e-12 Length: 462
 Score: 236.50 Matches: 54
 Percent Similarity: 49.68% Conservative: 23
 Best Local Similarity: 34.84% Mismatches: 49
 Query Match: 15.30% Indels: 29
 DB: 24 Gaps: 4

US-10-054-313-1 (1-286) x ABK64944 (1-462)

QY 140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgGlyProArgAlaGly 159
 DB 13 ATTGAATTTTACTGATGATCTTCTTAGGTAAT-----CCAGGGCGGGC 60
 QY 160 IleGlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArg 179
 DB 61 -----GGAATTGGTGCCGTATTGCTTATAAA 87
 QY 180 Gln-----ThrAsnGlnArgAlaGluIle 187
 DB 88 CAACATGAAAAAACAACCTCTCAAGGCTATTTCACCAACCCATATATCGAATGGAATTA 147
 QY 188 HisAlaAlaCysLysAlaIleGluAlaLysThrGlnAsnIleAsnLysLeuValLeu 207
 DB 148 CGCGCTGCTCATTTGAAGCAATTAATACATTAAGAACCTTGTCTG-----ATCAGCCTT 201
 QY 208 TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLys 227

| | | | |
|----|-----|---|-----|
| Db | 202 | TATAGTGATAGCCATATATGAAATAATGGCATACCAAAATGGATCTTTAACTGGAAAAA | 261 |
| Qy | 228 | AsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeu | 247 |
| Db | 262 | AluTaNTGGAAAGCAAGTCTGGAAAGCGCTGTAAAAACCACAGATTATGGATAGCGTTA | 321 |
| Qy | 248 | GluArgLeuThrGlnGlyMetAspIleGlnTrpMethIstValProGlyHisSerGlyPhe | 267 |
| Db | 322 | GATGAATCCATCCACGTCATATAATTAATGGCAATGGGTAAAAAGGCCATGCTGGACAC | 381 |
| Qy | 268 | IleGlyAsnGluAlaAspArgLeuAlaargGluGlyAlaLys | 282 |
| Db | 382 | AGAGAAAAATGAAATTTGGCATGAATTAGCAAAAAAAGGGCCAGAA | 426 |

Search completed: December 17, 2003, 19:23:34
Job time : 387 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 17, 2003, 19:11:56 ; Search time 73 Seconds

(without alignments)
1729.254 Million cell updates/sec

Title: US-10-054-313-1

Perfect score: 1546

Sequence: 1 MSWLLFLAHRVALAALPCRR.....FIGNEADRLAREGAKQSED 286

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -MODEL=frame+ p2n.model -DEV=xlh

-O=/cgn2_1/uspto.spool/US10054313/runat 17122003 150744 24207/app query.fasta_1.455
-DB=Issued Patents NA -QWTF=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054313 @CGN 1 1 56 @runat 17122003 150744 24207 -NCFU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----|----------------------|
| 1 | 241 | 15.6 | 1275 | 4 | US-09-252-991A-13638 |
| 2 | 236.5 | 15.3 | 1830121 | 4 | US-09-557-894-1 |
| 3 | 236.5 | 15.3 | 1830121 | 4 | US-09-643-980A-1 |
| 4 | 235.5 | 15.2 | 801 | 4 | US-09-199-637A-154 |
| C | 235.5 | 15.2 | 2048 | 4 | US-09-199-637A-136 |
| C | 233.5 | 15.1 | 570 | 4 | US-09-252-991A-13802 |
| 7 | 217 | 14.0 | 1509 | 4 | US-09-328-352-1765 |
| 8 | 174 | 11.3 | 501 | 1 | US-08-163-181-3 |
| 9 | 174 | 11.3 | 501 | 1 | US-08-465-161-3 |
| 10 | 174 | 11.3 | 552 | 1 | US-08-163-181-4 |
| 11 | 174 | 11.3 | 552 | 1 | US-08-465-161-4 |
| 12 | 160 | 10.3 | 8088 | 4 | US-09-315-127-4 |

| | | | | | | | |
|----|-------|------|------|-------|-------------------|--------------------|-------------------|
| 13 | 160 | 10.3 | 8535 | 3 | US-08-716-351A-1 | Sequence 1, Appli | |
| 14 | 157 | 10.2 | 8060 | 3 | US-08-766-528-1 | Sequence 1, Appli | |
| 15 | 156 | 10.1 | 8202 | 1 | US-08-258-420-13 | Sequence 13, Appli | |
| C | 154 | 10.0 | 3612 | 4 | US-09-265-013-3 | Sequence 3, Appli | |
| 16 | 154 | 10.0 | 7308 | 3 | US-09-011-745-3 | Sequence 3, Appli | |
| 17 | 154 | 10.0 | 7308 | 3 | US-09-011-745-4 | Sequence 4, Appli | |
| 18 | 154 | 10.0 | 7308 | 3 | US-09-011-745-5 | Sequence 5, Appli | |
| 19 | 154 | 10.0 | 7616 | 3 | US-09-011-745-2 | Sequence 2, Appli | |
| 20 | 154 | 10.0 | 8332 | 3 | US-08-850-961-1 | Sequence 1, Appli | |
| 21 | 154 | 10.0 | 8332 | 4 | US-09-479-776-1 | Sequence 1, Appli | |
| 22 | 154 | 10.0 | 8332 | 4 | US-09-309-572-11 | Sequence 11, Appli | |
| 23 | 154 | 10.0 | 8332 | 4 | US-09-315-127-1 | Sequence 1, Appli | |
| 24 | 154 | 10.0 | 8332 | 4 | US-09-265-013-1 | Sequence 1, Appli | |
| 25 | 154 | 10.0 | 8332 | 4 | US-09-554-572-25 | Sequence 25, Appli | |
| 26 | 153 | 9.9 | 6363 | 2 | US-08-929-967-6 | Sequence 6, Appli | |
| 27 | 152 | 9.8 | 8323 | 1 | US-08-110-300A-8 | Sequence 8, Appli | |
| 28 | 152 | 9.8 | 8323 | 2 | US-08-886-642-8 | Sequence 2, Appli | |
| 29 | 152 | 9.8 | 8323 | 4 | US-09-433-322B-2 | Sequence 2, Appli | |
| 30 | 152 | 9.8 | 8323 | 5 | PCT-US93-08041-8 | Sequence 8, Appli | |
| C | 31 | 152 | 9.8 | 10367 | 1 | US-08-110-300A-9 | Sequence 9, Appli |
| C | 32 | 152 | 9.8 | 10367 | 2 | US-08-886-642-9 | Sequence 9, Appli |
| C | 33 | 152 | 9.8 | 10367 | 5 | PCT-US93-08041-9 | Sequence 9, Appli |
| 34 | 149 | 9.6 | 8132 | 3 | US-08-766-528-3 | Sequence 3, Appli | |
| 35 | 144 | 9.3 | 8655 | 3 | US-09-075-272-1 | Sequence 1, Appli | |
| 36 | 140.5 | 9.1 | 9472 | 1 | US-08-325-547-9 | Sequence 9, Appli | |
| 37 | 139.5 | 9.0 | 9468 | 1 | US-08-325-547-10 | Sequence 10, Appli | |
| 38 | 135.5 | 8.8 | 3093 | 4 | US-08-811-682-14 | Sequence 14, Appli | |
| 39 | 135.5 | 8.8 | 3841 | 4 | US-08-811-682-13 | Sequence 13, Appli | |
| 40 | 135.5 | 8.8 | 9471 | 4 | US-08-811-682-1 | Sequence 1, Appli | |
| 41 | 132.5 | 8.6 | 9633 | 6 | 5223423-1 | Patent No. 5223423 | |
| 42 | 127.5 | 8.2 | 2689 | 2 | US-08-876-546A-15 | Sequence 15, Appli | |
| 43 | 127.5 | 8.2 | 2689 | 3 | US-09-412-252-15 | Sequence 15, Appli | |
| 44 | 127.5 | 8.2 | 3168 | 2 | US-08-859-251-14 | Sequence 14, Appli | |
| 45 | 127.5 | 8.2 | 3168 | 3 | US-09-256-490-14 | Sequence 14, Appli | |

ALIGNMENTS

RESULT 1

US-09-252-991A-13638

; Sequence 13638, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 13638

; LENGTH: 1275

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-13638

Alignment Scores:
Pred. No.: 3,79e-17 Length: 1275
Score: 241.00 Matches: 92
Percent Similarity: 35.96% Conservative: 36
Best Local Similarity: 25.84% Mismatches: 126
Query Match: 15.59% Indels: 102
DB: 4 Gaps: 14

US-10-054-313-1 (1-286) x US-09-252-991A-13638 (1-1275)

Oy 9 HieArgValAlaLeuAlaLeuProCysArgGlySerArgGlyPheGlyMetPhe 28

Db 259 CACCGGCAGATTACGCGCG-----TGTCGCTGGGCCCGCGCGGTGGA 312

QY 29 TyrAlaValAlaArgGlyArgLysThrGlyValPheLeuThrTrpAsnGluCysArgAla 48
DB 313 CATCCCTCGAAGAGGCGC-----CTGGCCGCTCAGCAACATGCGCGGA 360
QY 49 GlnValAspArgPheProAlaAlaArg-----PheLysLysPheAlaThrGluAsp 65
DB 361 -----CGTGGTCTCTGCTGCAACAGCGCTGGATTCTGCTGCTCACCCTCT 411
QY 66 GluAlaTrpAla-----PheValArgLysSerLysAlaSerProGluValSerGluGlyHisGlu 84
DB 412 CTTGGCGAAGCGCGCTACCTGCTGCTCGGCGCGCCACCTGCTGCTGATCGGCATCAA 471
QY 85 AsnGlnHis-----GlyGln 89
DB 472 CCCATGGAGCTGTGGGCGATCCGTCATTATTTCGCGGGGATGCCITGCGCCAGCGCG 531
QY 90 GluSerGluAlaLysProGlyLysArgLysGluArgLysProLysAspGlyAspGlyHisGlu 109
DB 532 CTGCATTCCTCGCTCGCGGCTGCGACTGCTGCTCAACCTGCTGCTGCTGCTGCTGCTG 591
QY 110 SerAla----- 111
DB 592 AGCGGCTTCGGTGTATCGTCCGCGCTGCTGCTGCGCAGCTGCGCAATCGCGCTGGC 651
QY 112 GlnProTyrAlaLysHisMetLys----- 119
DB 652 TCGCTTGGAGCTCGGGCGACGCTGCGCAGTCTTGGCGCGCGCTTCTATCTATTGGT 711
QY 120 -----ProSerValGluProAlaProValSerArgAspThr----- 132
DB 712 GGCACGCAAGCTGTGTGGTGTGGGCTGCGCCGCTGCGCAGAGCAAGCGCGACCGCGCG 771
QY 133 -----PheSerTyrMetGly--- 137
DB 772 TCAGCTGTGGCCATCGCGTGGCGAAGTACGCGCGGAGATTCGGAATTTAGCATG 831
QY 138 -----AspPheValValTyrThrAspGlyCysCysSerSerAsnGlyArg 154
DB 832 ACAGATAAAGACAGATGATGATATACCGCGCGCTGCGCAGCGCAACCTGGCGCG 891
QY 155 LysProArgAlaGlyIleGlyValTyr-----TrpGlyProGly 167
DB 892 GCGCGTGGGGCGCTGTCTCTACAGAGCGCGCGAGAGCTTTGGGCG---GGC 948
QY 168 HisProLysAsnValGlyIleArgLysProGlyArgGlnThrAsnGlnArgAlaGluLeu 187
DB 949 GAGCCG-----GACACCACCAACACCGCATGGAATG 981
QY 188 HisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnLysLeuValLeu 207
DB 982 ATGGCGCGATCCAGCGCTGGCGGCGCTACAGCGTCTCTGCTCCGATCGCTGATC--- 1038
QY 208 TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLys 227
DB 1039 ---ACCGACTCGAATACGTGATGCGCGCATCACCGAATGTTGCGCACTGGAAGA 1095
QY 228 AsnGlyTrpLysSerAlaGlyLysGluValIleAsnLysLysAspPheValAlaLeu 247
DB 1096 CCGCGTGGAGACCGCGCAAGCGCGTCAAGATCCGACCTCTGGCAGGCGCTG 1155
QY 248 GluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe 267
DB 1156 GATGAGCAGTTCGCGCGCACCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1215
QY 268 IleGlyAsnGlnLysAlaAspArgLysAlaArgGluGlyAlaLysGln 283
DB 1216 CCGCGCAACAGCG 1263

RESULT 2

US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 650581
; GENERAL INFORMATION:

APPLICANT: Fleischmann et al.
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1930121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1
Alignment Scores:
Pred. No.: 8,37e-12 Length: 1930121
Score: 236.50 Matches: 54
Percent Similarity: 49.68% Conservative: 23
Best Local Similarity: 34.84% Mismatches: 49
Query Match: 15.30% Indels: 29
DB: 4 Gaps: 4
US-10-054-313-1 (1-286) x US-09-557-884-1 (1-1930121)
QY 140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly 159
DB 152742 ATTGAATTTTACTGATGATCTTCTGTTAGGTAAT-----CCAGGGCGCGC 152789
QY 160 IleGlyValTyrTrpGlyProGlyHisProLysAsnValGlyIleArgLysProGlyArg 179
DB 152790 -----GGAATTGTCGCTGCTGTTATATAA 152816
QY 180 Gln-----ThrAsnGlnArgAlaGluLeu 187
DB 152817 CAACATGAAAAACACTCTCAAAGGTATTTCCAAACCAACCAATATCGAATGGAATTA 152876
QY 188 HisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnLysLeuValLeu 207
DB 152877 CGCGCTGTCATTGAAGCATTAATACATTAAGAACCTTGTCTG-----ATCACGCTT 152930
QY 208 TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLys 227
DB 152931 TATAGTATAGCCCAATATATGAAAAATGGCATACCAAAATGGATCTTTAATCTGAAAAA 152990
QY 228 AsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeu 247
DB 152991 AATAATGGAAGCAAGTCTCTGGAAGCGCTGTAAACCAAGATTTTATGATAGCCTTA 153050

[illegible]

RESULT 3

US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850

```

US-09-643-990A-1
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
Alignment Scores:
Pred. No.: 8,378-12 Length: 1830121
Score: 236.50 Matches: 54
Percent Similarity: 49.68% Conservative: 23
Best Local Similarity: 34.84% Mismatches: 49
Query Match: 15.30% Indels: 29
DB: 4 Gaps: 4

```

US-10-054-313-1 (1-286) X US-09-643-990A-1 (1-1830121)

| Qy | 140 | valvalvalValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly | 159 |
|----|--------|---|--------|
| Db | 152742 | ATTGAAATTTTACATGATGATCTTGCTAGGTAAAT-----CCAGGGCGGGC | 152789 |

| | | | |
|----|--------|---|--------|
| Qy | 160 | IleGlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyAsp | 179 |
| Db | 152790 | -----GGAATTGGTGCCCTATTGGCTTTATAAA | 152816 |
| Qy | 180 | Gln-----ThrAsnGlnArgAlaGluIle | 187 |
| Db | 152817 | CAACATGAAAAACACTCTCCAAAGGCTATTTCAAACACCCCAATTAATCGAATGGAAATTA | 152876 |
| Qy | 188 | HisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeu | 207 |
| Db | 152877 | CCGCGTGTCAATGAAGCATTAATAATACATTAAAGAACCCTTGCTTG-----ATCACGCTT | 152930 |
| Qy | 208 | TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLys | 227 |
| Db | 152931 | TATAGTGATAGCCAATATATGAAAATATGGCATACCAAAATGSGATCTTTAACTGGAATAAAA | 152990 |
| Qy | 228 | AsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeu | 247 |
| Db | 152991 | AATAATTGGAAGCAGATCTGGAAAGCCGTGTAATAAACCAAGATTATGTGATAGGCCTTA | 153050 |
| Qy | 248 | GluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe | 267 |
| Db | 153051 | GATGAATCCATCCACAGCTCATAAATTAATTGGCAATGGGTAAAGGCCATGCTGGACAC | 153110 |
| Qy | 268 | IleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAlaLys | 282 |
| Db | 153111 | AGAGAAATGAAATTTGGCATGAATTTAGCAAAAAAGGGCGAGAA | 153155 |

RESULT 4

```

US-09-199-637A-154
; Sequence 154, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Tsengalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 154
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-154

```

| | | |
|------------------------|----------|------------------|
| Alignment Scores: | | |
| Pred. No.: | 7.55e-17 | 801 |
| Score: | 235.50 | 60 |
| Percent Similarity: | 49.12% | Conservative: 64 |
| Best Local Similarity: | 35.09% | Mismatches: 24 |
| Query Match: | 15.23% | Indels: 24 |
| DB: | 4 | Gaps: 5 |

US-10-054-313-1 (1-286) x US-09-199-637A-154 (1-801)

| | | | |
|----|-----|---|-----|
| Qy | 124 | ProAlaProProValSerArgAspThrPheSerTyrMetGly-----AspPhe | 139 |
| | | ::: | ::: |
| Db | 314 | CCGGTGCGCAAGTCAGCCGCG-AGATTCCGAATTTAGGCATGCACAGATAAAGAACAAG | 372 |
| Qy | 140 | ValValValTyrThrAspGlyCysCysserSerIasnGlyArgArgGlyysProArgAlaGly | 159 |
| | | ::: | |

```

Db 373 GTAGTGATCTATACCGACGGCGCTGCAAGGCAACCTGGGCGCGGGCTGGGGGCG 432
Qy 160 IleGlyValTyr-----TTPGlyProGlyHisProLeuAsnVal 172
Db 433 TTGCTCTCTATCAAGGCGCCGAGCAGAGCTTTGGGGC---GGCAGCGC-----480
Qy 173 GlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLys 192
Db 481 -----GACACCACCAACCGCATGGAGCTGATGGCGGCGATCCAG 522
Qy 193 AlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMet 212
Db 523 GGCTGGCGGCGACTCAAGCGTCTCTCCGATCCGCTCTGATC-----ACGACTCGGAA 576
Qy 213 PheThrIleAsnGlyIleThrAsnTrpValGlnGlyTyrPlyLysAsnGlyTyrLysThr 232
Db 577 TACGTGATGGCGGCATCACCGAATGTTGCCGAACCTGGAAGAGCGCGCTGGAAGACC 636
Qy 233 SerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGln 252
Db 637 GCAGCAAGCAGCGCTGCAAGATGCGGACCTCTGGCAGCGCCTGGATGAACAGGTCCGC 696
Qy 253 GlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGlu 272
Db 697 CGCACCAAGTGGAGTGGCAGTGGTCCGCGGCGCATACCGGCGACCCCGGCAACAGCGG 756
Qy 273 AlaAspArgLeuAlaArgGluGlyAlaLysGln 283
Db 757 GCCGACCAAGTGGCCAAACCGTGGCGCTGCCGCGAA 789

```

RESULT 5

US-09-199-637A-136/C
; Sequence 136, Application US/09199637A

; Patent No. 6355411

; GENERAL INFORMATION:

; APPLICANT: Ausubel, Frederick

; APPLICANT: Goodman, Howard M.

; APPLICANT: Rahme, Laurence G.

; APPLICANT: Mahajan-Wiklos, Shalina

; APPLICANT: Tan, Man-Wah

; APPLICANT: Cao, Hui

; APPLICANT: Drenkard, Eliana

; APPLICANT: Tsongalis, John

; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID

; FILE REFERENCE: 00786/361002

; CURRENT APPLICATION NUMBER: US/09/199,637A

; PRIOR FILING DATE: 1998-11-25

; PRIOR FILING DATE: 1997-11-25

; NUMBER OF SEQ ID NOS: 437

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 136

; LENGTH: 2048

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-199-637A-136

Alignment Scores:

| Pred. No.: | 3,19e-16 | Length: | 2048 |
|------------------------|----------|---------------|------|
| Score: | 235.50 | Matches: | 60 |
| Percent Similarity: | 49.12% | Conservative: | 24 |
| Best Local Similarity: | 35.09% | Mismatches: | 64 |
| Query Match: | 15.23% | Indels: | 24 |
| DB: | 4 | Gaps: | 5 |

US-10-054-313-1 (1-286) x US-09-199-637A-136 (1-2048)

```

Qy 124 ProAlaProValSerArgAspThrPheSerTyrMetGly-----AspPhe 139
Db 648 CCGGTGGCGAAGTCAAGCGCGCG-AGATTCCGAAATTTAGGCATGACAGATAAAGAACAG 590
Qy 140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly 159

```

```

Db 589 GTAGTGATCTATACCGACGGCGCTGCAAGGCAACCTGGGCGCGGGCTGGGGGCG 530
Qy 160 IleGlyValTyr-----TTPGlyProGlyHisProLeuAsnVal 172
Db 529 TTGCTCTCTATCAAGGCGCCGAGCAGAGCTTTGGGGC---GGCAGCGC-----482
Qy 173 GlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLys 192
Db 481 -----GACACCACCAACCGCATGGAGCTGATGGCGGCGATCCAG 440
Qy 193 AlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMet 212
Db 439 CGCTGGCGGCGACTCAAGCGTCTCTCCGATCCGCTCTGATC-----ACGACTCGGAA 386
Qy 213 PheThrIleAsnGlyIleThrAsnTrpValGlnGlyTyrPlyLysAsnGlyTyrLysThr 232
Db 385 TACGTGATGGCGGCATCACCGAATGTTGCCGAACCTGGAAGAGCGCGCTGGAAGACC 326
Qy 233 SerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGln 252
Db 325 GCCAGCAAGCAGCGCTGCAAGATGCGGACCTCTGGCAGCGCCTGGATGAACAGGTCCGC 266
Qy 253 GlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGlu 272
Db 265 CGCACCAAGTGGAGTGGCAGTGGTCCGCGGCGCATACCGGCGACCCCGGCAACAGCGG 206
Qy 273 AlaAspArgLeuAlaArgGluGlyAlaLysGln 283
Db 205 GCCGACCAAGTGGCCAAACCGTGGCGCTGCCGCGAA 173

```

RESULT 6

US-09-252-991A-13802/C

; Sequence 13802, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 13802

; LENGTH: 570

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-13802

Alignment Scores:

| Pred. No.: | 7,46e-17 | Length: | 570 |
|------------------------|----------|---------------|-----|
| Score: | 233.50 | Matches: | 54 |
| Percent Similarity: | 50.33% | Conservative: | 22 |
| Best Local Similarity: | 35.76% | Mismatches: | 56 |
| Query Match: | 15.10% | Indels: | 19 |
| DB: | 4 | Gaps: | 4 |

US-10-054-313-1 (1-286) x US-09-252-991A-13802 (1-570)

```

Qy 140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly 159
Db 546 GTAGTGATCTATACCGACGGCGCTGCAAGGCAACCTGGGCGCGGGCTGGGGGCG 487
Qy 160 IleGlyValTyr-----TTPGlyProGlyHisProLeuAsnVal 172
Db 486 TTGCTCTCTATCAAGGCGCCGAGCAGAGCTTTGGGGC---GGCAGCGC-----439
Qy 173 GlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLys 192

```



```

; MOLECULE TYPE: DNA (genomic)
; US-08-465-161-3
Alignment Scores:
Pred. No.: 2,41e-10 Length: 501
Score: 174.00 Matches: 51
Percent Similarity: 51.3% Conservative: 25
Best local Similarity: 34.4% Mismatches: 59
Query Match: 11.2% Indels: 14
DB: 1 Gaps: 5

US-10-054-313-1 (1-286) x US-08-465-161-3 (1-501)
QY 140 ValValValValTyrThrAspGlyCysCysSerSerAsnGlyValArgLysProArgAlaGly 159
Db 25 GTGGCCCTTTTACCACCGGGGCTCTCTGGGAAACCC-CGGGCCCGG-----GGG 74
QY 160 IleGlyValTyrTrpGlyPro-----GlyHisProLeuAsnValGlyIleArgLeu 176
Db 75 GTGGGC-----GGCCCTCTCTCGCTTCCACGCCACGAGAGCTCTCTCCGGGG 125
QY 177 ProGlyArgGln-ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGl 196
Db 126 AGAGGGCTGCACCAACCAACCGCATGAGCTCAAGGGCGGCCTAGAGGGCCTAAAGGC 185
QY 196 nAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAs 216
Db 186 CCTCAGGAGCCT-----TGGAGGTGGACCTCTACCCGACGACCTACCTCAAGAA 239
QY 216 nGlyIleThr---AsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGl 235
Db 240 GGCCTTCACCGAGGCTGGCTGGAGGCTGGCGGAAAGGGCTGGCGGACGGCGAGGG 299
QY 235 YLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAs 255
Db 300 CAAGCCCGTGAAACCGGACCTCTGGGAGGCGCTCTCTCGCCATGGCCCCCACC 359
QY 255 pIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluAlaAspAr 275
Db 360 GTGGCGCTTCCACTTCGTGAAGGGGCACACGGGCCACCGGGAGAACGAACGGGTGACCG 419
QY 275 gLeuAlaArgGluGlyAlaLys 282
Db 420 GGAGGCGAGGCGCCAGGCCCG 441

RESULT 9
; Sequence 3, Application US/08465161
; Patent No. 5500370
; GENERAL INFORMATION:
; APPLICANT: Jendrisak, Jerome J.
; APPLICANT: Smith, Robert E.
; APPLICANT: Dahl, Gary A.
; TITLE OF INVENTION: THERMOSTABLE RIBONUCLEASE H AND GENETIC CONSTRUCTS THEREFORE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,161
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/815,095
; FILING DATE: 27-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: P-35,433
; REFERENCE/DOCKET NUMBER: 31-307-9001-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 277-5774
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 501 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: DNA (genomic)
; US-08-465-161-3
Alignment Scores:
Pred. No.: 2,41e-10 Length: 501
Score: 174.00 Matches: 51
Percent Similarity: 51.3% Conservative: 25
Best local Similarity: 34.4% Mismatches: 59
Query Match: 11.2% Indels: 14
DB: 1 Gaps: 5

US-10-054-313-1 (1-286) x US-08-465-161-3 (1-501)
QY 140 ValValValValTyrThrAspGlyCysCysSerSerAsnGlyValArgLysProArgAlaGly 159
Db 25 GTGGCCCTTTTACCACCGGGGCTCTCTGGGAAACCC-CGGGCCCGG-----GGG 74
QY 160 IleGlyValTyrTrpGlyPro-----GlyHisProLeuAsnValGlyIleArgLeu 176
Db 75 GTGGGC-----GGCCCTCTCTCGCTTCCACGCCACGAGAGCTCTCTCCGGGG 125
QY 177 ProGlyArgGln-ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGl 196
Db 126 AGAGGGCTGCACCAACCAACCGCATGAGCTCAAGGGCGGCCTAGAGGGCCTAAAGGC 185
QY 196 nAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAs 216
Db 186 CCTCAGGAGCCT-----TGGAGGTGGACCTCTACCCGACGACCTACCTCAAGAA 239
QY 216 nGlyIleThr---AsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGl 235
Db 240 GGCCTTCACCGAGGCTGGCTGGAGGCTGGCGGAAAGGGCTGGCGGACGGCGAGGG 299
QY 235 YLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAs 255
Db 300 CAAGCCCGTGAAACCGGACCTCTGGGAGGCGCTCTCTCGCCATGGCCCCCACC 359
QY 255 pIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluAlaAspAr 275
Db 360 GTGGCGCTTCCACTTCGTGAAGGGGCACACGGGCCACCGGGAGAACGAACGGGTGACCG 419
QY 275 gLeuAlaArgGluGlyAlaLys 282
Db 420 GGAGGCGAGGCGCCAGGCCCG 441

RESULT 10
; US-08-163-181-4
; Sequence 4, Application US/08163181
; Patent No. 5459055
; GENERAL INFORMATION:
; APPLICANT: Jendrisak, Jerome J.
; APPLICANT: Smith, Robert E.
; APPLICANT: Dahl, Gary A.
; TITLE OF INVENTION: THERMOSTABLE RIBONUCLEASE H
; TITLE OF INVENTION: AND GENETIC CONSTRUCT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/163,181
; FILING DATE:

```



```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/815,095
;   FILING DATE: 27-DEC-1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Baker, Jean C.
;     REGISTRATION NUMBER: P-35,433
;     REFERENCE/DOCKET NUMBER: 31-307-9001-1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (414) 277-5709
;     TELEFAX: (414) 277-5774
;   INFORMATION FOR SEQ ID NO: 4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 552 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: double
;       TOPOLOGY: linear
;   MOLECULE TYPE: Other Nucleic Acid
; US-08-163-181-4
;
; Alignment Scores:
; Pred. No.: 2,79e-10 Length: 552
; Score: 174.00 Matches: 51
; Percent Similarity: 51.35% Conservatives: 25
; Best Local Similarity: 34.46% Mismatches: 59
; Query Match: 11.25% Indels: 14
; DB: 1 Gaps: 5
;
; US-10-054-313-1 (1-286) x US-08-163-181-4 (1-552)
;
; Qy 140 ValValValValThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGly 159
;   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; Db 76 GTGGCCCTCTTCCACCGCGGCGCTGCTGGGAAACCC-CGGGCCCGG-----GGG 125
;
; Qy 160 IleGlyValTyTrpGlyPro-----GlyHisProLeuAsnValGlyIleArgLeu 176
;   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; Db 126 GTGGGC-----GGCCCTCTCCGCTTCCACGCCACGAGAGCTCTCTCCGGGG 176
;
; Qy 177 ProGlyArgGln-ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGl 196
;   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; Db 177 AGAGGCGCTGCACCAACCAACCGCATGAGCTCAAGCGCGGCATAGAGGCGCTAAAGGC 236
;
; Qy 196 nAlaLysThrGlnAsnIleAsnLysLeuValLeuTyThrAspSerMetPheThrIleAs 216
;   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; Db 237 CCTCAGGAGCCT-----TGCAGGTGGACCTCTACCCAGCAGCAGCCTACTCTCAAGAA 290
;
; Qy 216 nGlyIleThr---AsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGl 235
;   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; Db 291 GGCCTTCCAGGCGCTGGCTGGAAGGCTGGCGGAAAGGGGTGGCGGAGGCGGAGGG 350
;
; Qy 235 YLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAs 255
;   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; Db 351 CAAGCCCGTGAAACCGGACCTCTGGAGGCGCTCTCTCGGCATGGCCCCCAGCG 410
;
; Qy 255 pIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspAr 275
;   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; Db 411 GGTGGCTTCCACTCTCGTGAAGGGGCACACGGGCCACCGGAGAACGCGGTGGACCG 470
;
; Qy 275 GLeuAlaArgGluGlyValAlaLys 282
;   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; Db 471 GGAGGCGAGGCGGCCAGGCCAG 492
;
; RESULT 11
; US-08-465-161-4
; Sequence 4, Application US/08465161
; Patent No. 5500370
; GENERAL INFORMATION:
; APPLICANT: Jendrisak, Jerome J.
; APPLICANT: Smith, Robert E.
; APPLICANT: Dahl, Gary A.
; TITLE OF INVENTION: THERMOSTABLE RIBONUCLEASE H AND GENETIC CONSTRUCTS THEREFORE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/465,161
;     FILING DATE:
;     CLASSIFICATION: 435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 07/815,095
;       FILING DATE: 27-DEC-1991
;       ATTORNEY/AGENT INFORMATION:
;         NAME: Baker, Jean C.
;         REGISTRATION NUMBER: P-35,433
;         REFERENCE/DOCKET NUMBER: 31-307-9001-1
;       TELECOMMUNICATION INFORMATION:
;         TELEPHONE: (414) 277-5709
;         TELEFAX: (414) 277-5774
;       INFORMATION FOR SEQ ID NO: 4:
;         SEQUENCE CHARACTERISTICS:
;           LENGTH: 552 base pairs
;           TYPE: nucleic acid
;           STRANDEDNESS: double
;           TOPOLOGY: linear
;         MOLECULE TYPE: Other Nucleic Acid
;       US-08-465-161-4
;
; Alignment Scores:
; Pred. No.: 2,79e-10 Length: 552
; Score: 174.00 Matches: 51
; Percent Similarity: 51.35% Conservatives: 25
; Best Local Similarity: 34.46% Mismatches: 59
; Query Match: 11.25% Indels: 14
; DB: 1 Gaps: 5
;
; US-10-054-313-1 (1-286) x US-08-465-161-4 (1-552)
;
; Qy 140 ValValValValThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGly 159
;   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; Db 76 GTGGCCCTCTTCCACCGCGGCGCTGCTGGGAAACCC-CGGGCCCGG-----GGG 125
;
; Qy 160 IleGlyValTyTrpGlyPro-----GlyHisProLeuAsnValGlyIleArgLeu 176
;   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; Db 126 GTGGGC-----GGCCCTCTCCGCTTCCACGCCACGAGAGCTCTCTCCGGGG 176
;
; Qy 177 ProGlyArgGln-ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGl 196
;   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; Db 177 AGAGGCGCTGCACCAACCAACCGCATGAGCTCAAGCGCGGCATAGAGGCGCTAAAGGC 236
;
; Qy 196 nAlaLysThrGlnAsnIleAsnLysLeuValLeuTyThrAspSerMetPheThrIleAs 216
;   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; Db 237 CCTCAGGAGCCT-----TGCAGGTGGACCTCTACCCAGCAGCAGCCTACTCTCAAGAA 290
;
; Qy 216 nGlyIleThr---AsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGl 235
;   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; Db 291 GGCCTTCCAGGCGCTGGCTGGAAGGCTGGCGGAAAGGGGTGGCGGAGGCGGAGGG 350
;
; Qy 235 YLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAs 255
;   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; Db 351 CAAGCCCGTGAAACCGGACCTCTGGAGGCGCTCTCTCGGCATGGCCCCCAGCG 410
;
; Qy 255 pIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspAr 275
;   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; Db 411 GGTGGCTTCCACTCTCGTGAAGGGGCACACGGGCCACCGGAGAACGCGGTGGACCG 470

```

QY 275 gLeuAlaArgGluGlyAlaLys 282
 DB 471 GGAGGCGAGGCCAGGCCAG 492
 RESULT 12
 US-09-315-127-4
 ; Sequence 4, Application US/09315127
 ; Patent No. 6448390
 ; GENERAL INFORMATION:
 ; APPLICANT: The University of Tennessee, c/o Richard Cox
 ; TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and
 ; FILE REFERENCE: 44137-5023, U. of Tennessee
 ; CURRENT APPLICATION NUMBER: US/09/315,127
 ; CURRENT FILING DATE: 1999-05-20
 ; NUMBER OF SEQ ID NOS: 23
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 4
 ; LENGTH: 8088
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: SEQ. ID NO. 3,
 ; OTHER INFORMATION: retroviral vector
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (5552)..(7552)
 US-09-315-127-4
 Alignment Scores:
 Pred. No.: 6,13e-07 Length: 8088
 Score: 160.00 Matches: 62
 Percent Similarity: 43.00% Conservative: 27
 Best Local Similarity: 29.95% Mismatches: 70
 Query Match: 10.35% Indels: 48
 DB: 4 Gaps: 10
 US-10-054-313-1 (1-286) x US-09-315-127-4 (1-8088)
 QY 90 GluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyHisGlu 109
 DB 4000 GAGTCGAGGCCAGGCCAGTGCACAGTGTCTCAGAAATCCTC----- 4041
 QY 110 SerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProValSer 129
 DB 4042 ---GCCGAAGAACTGGAACTCGACGAGACCTAGAACACCACTTCCCGGGGTGCA 4098
 QY 130 ArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCysCys--- 148
 DB 4099 -----ACCTGGTATACAGACGCTAGCAGTTTC 4125
 QY 149 SerSerAsnGlyArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGlyHis 168
 DB 4126 ATCAGGAAGGTAACGGAGA-----GCAGGGGCC 4155
 QY 169 ProLeuAsnValGlyIleArg-----LeuProGlyArgGlnThrAsn 182
 DB 4156 CCATCGTAGTCGAGCGAGCGGATGGGTAGCAGCGCTGCCAGAGTACGTCAGCC 4215
 QY 183 GlnArgAlaGluIleHisAlaLysLysAlaIleGluGlnAlaLysThrGlnAsnIle 202
 DB 4216 CAGAGGCTGAACCTAGTACCTTTGACGAGCAGCATACGCTGCCGCAAGAAACATC 4275
 QY 203 AsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpVal 222
 DB 4276 AAC-----ATCTACAGGACGAGCATGATGCTTTT-----GCCACTGCTCATATT 4320
 QY 223 GlnGly-----TrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
 DB 4321 CATGGGCAATATATAAGCAGAGGGGCTCTCTCTCTCTCTGAAAGATATCAAAAC 4380
 QY 241 LysGluAspPheValIle-----LeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMet 259
 DB 4469 -----ACCTGGTATACAGAGCGGTAGCAGTTTC 4495

DB 4381 AAAGAGAAATTTGGCCCTGCTAGAGGCCATCCATCTCCCTAGGCGGTGCCATTATC 4440
 QY 260 HisValProGlyHisSerGly-----PheIleGlyAsnGluGluAlaAsp 274
 DB 4441 CACTGTCTGGCCACCGAGGAGGAAGTAACCTGTGGCCATCTGGGAACCGAGGGCCGAC 4500
 QY 275 ArgLeuAlaArgGluGlyAla 281
 DB 4501 GAGGCTGCAAGCAAGCCGCC 4521
 RESULT 13
 US-08-716-351A-1
 ; Sequence 1, Application US/08716351A
 ; Patent No. 6033905
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: Gibbon Ape Leukemia Virus-Based
 ; TITLE OF INVENTION: Retroviral Vectors
 ; NUMBER OF SEQUENCES: 5
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/716,351A
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US94/03784
 ; FILING DATE: 06-APR-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Bastian, Kevin L.
 ; REGISTRATION NUMBER: 34,774
 ; REFERENCE/DOCKET NUMBER: 15280-128-1PC
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 543-9600
 ; TELEFAX: (415) 543-5043
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 8535 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: 1..8535
 ; OTHER INFORMATION: /standard_name= "GalV SEATO Genome"
 US-08-716-351A-1
 Alignment Scores:
 Pred. No.: 6,66e-07 Length: 8535
 Score: 160.00 Matches: 62
 Percent Similarity: 43.00% Conservative: 27
 Best Local Similarity: 29.95% Mismatches: 70
 Query Match: 10.35% Indels: 48
 DB: 3 Gaps: 10
 US-10-054-313-1 (1-286) x US-08-716-351A-1 (1-8535)
 QY 90 GluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyHisGlu 109
 DB 4370 GAGTCGAGGCCAGGCCAGTGCACAGTGTCTCAGAAATCCTC----- 4411
 QY 110 SerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProValSer 129
 DB 4412 ---GCCGAAGAACTGGAACTCGACGAGACCTAGAACACCACTTCCCGGGGTGCA 4468
 QY 130 ArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCysCys--- 148
 DB 4469 -----ACCTGGTATACAGAGCGGTAGCAGTTTC 4495

QY 149 SerSerAsnGlyArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGlyHis 168
 Db 4496 ATCAGGAAGGTAAACGGAGA-----GCAGGGGCC 4525
 QY 169 ProLeuAsnValGlyIleArg-----LeuProGlyArgGlnThrAsn 182
 Db 4526 CCGATCTAGATGGCAACGGACGGATGGCTAGCAGCTGCCAGAGGTACCTCAGCC 4585
 QY 183 GlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIle 202
 Db 4586 CAGAAGGCTGAACCTAGTACCTTGCAGCGACGATTCAGCCTGGCGGAAGGAACATC 4645
 QY 203 AsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpVal 222
 Db 4646 AAC-----ATCTACCGACAGCAGGATGCTTTT-----GCCACTGCTCATATT 4690
 QY 223 GlnGly-----TrpLysIleAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
 Db 4691 CATGGGGCAATATATAACAGAGGGGGTGTCTCCTCTGTGAAAAGATATCAAAAC 4750
 QY 241 LysGluAspPheValAla---LeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMet 259
 Db 4751 AAAGAGGAAATTTTGGCCCTGCTAGAGGCCATCCATCTCCCTAGCGGGTGCCTATTC 4810
 QY 260 HisValProGlyHisSerGly-----PheIleGlyAsnGluGluAlaAsp 274
 Db 4811 CACTCTCTGGCCACAGAGGGAAGTAACCTGTGGCCACTGGGAACCGGAGGCCGAC 4870
 QY 275 ArgLeuAlaArgGluGlyAla 281
 Db 4871 GAGGCTGCAGAACAGCGGCC 4891

RESULT 14

US-08-766-528-1
 ; Sequence 1, Application US/08766528
 ; Patent No. 6190861

; GENERAL INFORMATION:
 ; APPLICANT: Jav A. Fishman
 ; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
 ; TITLE OF INVENTION: AND METHODS OF USE

; NUMBER OF SEQUENCES: 74
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD, LLP
 ; STREET: 60 State Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109-1875

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/766,528

; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/572,645
 ; FILING DATE: 14-DEC-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Louis Myers
 ; REGISTRATION NUMBER: 35,965
 ; REFERENCE/DOCKET NUMBER: MGP-038CP
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)227-7400
 ; TELEFAX: (617)227-5941

; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 8060 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear

; MOLECULE TYPE: cDNA
 ; US-08-766-528-1

Alignment Scores:
 Pred. No.: 1,31e-06 Length: 8060
 Score: 157.00 Matches: 52
 Percent Similarity: 50.31% Conservative: 29
 Best Local Similarity: 32.30% Mismatches: 50
 Query Match: 10.16% Indels: 30
 DB: 3 Gaps: 9

US-10-054-313-1 (1-286) x US-08-766-528-1 (1-8060)

QY 137 GlyAspPheValValTyrThrAspGlyCysCysSerSer-----AsnGlyArg 153
 Db 6585 GGAGAAGTCTTAACCTGGTTCACTACGGA-----AGCAGCTATGTGGTGAAGGTAAG 6638
 QY 154 ArgLysProArgAla-----GlyIleGlyValTyrTrpGlyProGlyHisPro 169
 Db 6639 AGGATGGCTGGGGCGGGTGGACGGACCCGACCATCTGGGCCAGCAGC----- 6692
 QY 170 LeuAsnValGlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAla 189
 Db 6693 -----CTGCGGGAGGAACTTCAGCACAAAAGGCTCAGCTCATGGCC 6734
 QY 190 AlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThr 209
 Db 6735 CTCACCAAGCTTTGGCGCTGGCGGAGGAAATCCATAAC-----ATTATACG 6785
 QY 210 AspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGly-----TrpLysLys 227
 Db 6786 GACAGCAGGTATGCCCTTT-----GGGACTGCACACGTACATGGGGCCATCTATAAACAA 6839
 QY 228 AsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAla--- 246
 Db 6840 AGGGGGTTGCTTACCTCAGCAGGAGGAAATAAAGAACAAAGAGGAAATTTCTAAGCCTA 6899
 QY 247 LeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGly 266
 Db 6900 TTAGAAGCCGTACATTTACCAAAAAGGCTAGCTATTATACACTGCTCTGCACATCAGAAA 6959
 QY 267 -----PheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAla 281
 Db 6960 GCTAAAGATCTCATATCCAGAGAAACCAAGATGGCTGACCGGGTTGCCAAGCAGCAGCC 7019
 QY 282 Lys 282
 Db 7020 CAG 7022

RESULT 15

US-08-258-420-13/c
 ; Sequence 13, Application US/08258420
 ; Patent No. 5710037

; GENERAL INFORMATION:

; APPLICANT: Nienhuis, Arthur W.

; APPLICANT: Vanin, Ello F.

; TITLE OF INVENTION: No. 5710037el Retroviral Envelope and LTR and Retroviral Vect.

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein

; STREET: 6 Becker Farm Road

; CITY: Roseland

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07068

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch diskette

; COMPUTER: IBM PS/2

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: WordPerfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/258,420

; FILING DATE: 10-JUN-1994

```
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Olatein, Elliot M.
/ REGISTRATION NUMBER: 24,025
/ REFERENCE/DOCKET NUMBER: 271010-208
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 201-994-1700
/ TELEFAX: 201-994-1744
/ INFORMATION FOR SEQ ID NO: 13:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 8202 bases
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: genomic DNA
/ FEATURE:
/ NAME/KEY: viral genome
US-08-258-420-13
```

```
Alignment Scores:
Pred. No.: 1.74e-06 Length: 8202
Score: 156.00 Matches: 65
Percent Similarity: 41.30% Conservative: 30
Best Local Similarity: 28.26% Mismatches: 68
Query Match: 10.09% Indels: 67
DB: 1 Gaps: 12
```

US-10-054-313-1 (1-286) x US-08-258-420-13 (1-8202)

```
QY 95 ProGlyLysArgLeuArgGluProLeuAspGlyAspGly-----HisGluSer 110
DB 3805 CCGGATGACCCATTACCAAGCCCTGCTCTGGACACGCGGTCACGTTCCGGCCAGT 3746
QY 111 AlaGlnProTyAlaLysHisMetLysProSerVal----- 122
DB 3745 AGTGGCCCTAAATCCAGCTACGCTGCTCCCTCTGCTGAGGAGGGGCTGCACATGACTG 3686
QY 123 -----GluProAlaProPro 127
DB 3685 CTTGACATCTTGCTGAAGCCACGGAACCTAGATCAGATCTACGGACCGCCCTCC 3626
QY 128 ValSerArg-AspThrPheSerTyrrMetGlyAspPheValValTyrrAspGlyCy 147
DB 3625 AGACGCGCACACACCTGG-----TACACGGATGGGAG 3593
QY 147 sCys---SerSerAsnGlyArgGlyProArgAlaGlyIle-----GlyVa 162
DB 3592 CAGCTTCTCAAGAGGGGAGGTAAGCGCGGAGCAGCGGTACCACTGAGACTGAGT 3533
QY 162 lTyrrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrAs 182
DB 3532 AATCTGGCCAGGGCA-----TTGCCACGGGACATCGGC 3497
QY 182 nGlnArgAlaGluIleHisAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnI 202
DB 3496 CCAAGAGCTGAATGATAGCGTCCCAAGCCCTAAAGATGGCAGAGGTAAGAGCT 3437
QY 202 eAsnLysLeuValLeuTyrrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpVa 222
DB 3436 AAAT-----GTTTATACTAGTCGCTTACGCTTTT-----GCCACGCCCATAT 3392
QY 222 lGlnGly-----TrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAs 240
DB 3391 TCATGGGAATATACAGAGGCGCGGTTGCTCATATCAGAGGAAAGAGATCAAGAA 3332
QY 240 nLysGluAspPheValAlaLeuGluArg-----LeuThrGlnGlyMetAspIleG 257
DB 3331 CAAGGACGAGNTCTAGCCCTACTAAGGCTCTCTTCTGCCCAAGAGACTTAGCATA-- 3274
QY 257 nTrpMetHisValProGlyHis-----SerGlyPheIleGlyAsnGluG 272
```

```
Db 3273 ----ATTCATTGCCCGGACATCAAAAAGGAACACGCGAGAGGCCAGGGCAACCGGAT 3218
QY 272 uAlaAspArgLeuAlaArgGluGlyAla 281
DB 3217 GGCCGACCAAGCGGCCCGAGAGTAGCC 3190
```

Search completed: December 17, 2003, 21:29:24
Job time : 1091 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 17, 2003, 20:35:09 ; Search time 365 Seconds

(without alignments)
2610.987 Million cell updates/sec

Title: US-10-054-313-1

Perfect score: 1546

Sequence: 1 MSWLLFLAHRVALAALPCRR.....FIGNEEDRLAREGAQKQSED 286

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2211978 seqs, 1666101734 residues

Total number of hits satisfying chosen parameters: 4423956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US10054313/runat 17122003 150747 24333/app query.fasta 1.455
-DB=Published Applications NA -OPMT=fastCap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10054313 @CGN 1.1 353 @runat 17122003 150747 24333
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FCGAPOP=6 -FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

- 1: /cgn2.6/ptodata/1/pubna/US07_PUBCOMB.seq*
- 2: /cgn2.6/ptodata/1/pubna/FCR_NEW_PUB.seq*
- 3: /cgn2.6/ptodata/1/pubna/US06_NEW_PUB.seq*
- 4: /cgn2.6/ptodata/1/pubna/US06_PUBCOMB.seq*
- 5: /cgn2.6/ptodata/1/pubna/US07_NEW_PUB.seq*
- 6: /cgn2.6/ptodata/1/pubna/PCTUS_PUBCOMB.seq*
- 7: /cgn2.6/ptodata/1/pubna/US08_NEW_PUB.seq*
- 8: /cgn2.6/ptodata/1/pubna/US08_PUBCOMB.seq*
- 9: /cgn2.6/ptodata/1/pubna/US09A_PUBCOMB.seq*
- 10: /cgn2.6/ptodata/1/pubna/US09B_PUBCOMB.seq*
- 11: /cgn2.6/ptodata/1/pubna/US09C_PUBCOMB.seq*
- 12: /cgn2.6/ptodata/1/pubna/US09_NEW_PUB.seq*
- 13: /cgn2.6/ptodata/1/pubna/US09_NEW_PUB.seq2*
- 14: /cgn2.6/ptodata/1/pubna/US10A_PUBCOMB.seq*
- 15: /cgn2.6/ptodata/1/pubna/US10B_PUBCOMB.seq*
- 16: /cgn2.6/ptodata/1/pubna/US10_NEW_PUB.seq*
- 17: /cgn2.6/ptodata/1/pubna/US60_NEW_PUB.seq*
- 18: /cgn2.6/ptodata/1/pubna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match Length | ID | Description |
|------------|-------|--------------|----|-------------|
|------------|-------|--------------|----|-------------|

RESULT 1
US-10-198-846-12889
; Sequence 12889, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillic, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steimann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MEI-049
; CURRENT APPLICATION NUMBER: US/10198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 12889
; LENGTH: 1807
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

Sequence 12889, A
Sequence 9423, Ap
Sequence 9423, Ap
Sequence 7011, Ap
Sequence 7011, Ap
Sequence 29506, A
Sequence 19582, A
Sequence 6980, Ap
Sequence 3900, Ap
Sequence 431, App
Sequence 13698, A
Sequence 943, App
Sequence 7313, Ap
Sequence 71, Appli
Sequence 1, Appli
Sequence 154, App
Sequence 136, App
Sequence 1, Appli
Sequence 1127, Ap
Sequence 1, Appli
Sequence 2, Appli
Sequence 62, Appl
Sequence 85, Appl
Sequence 1, Appli
Sequence 1, Appli
Sequence 15, Appl
Sequence 17, Appl
Sequence 1, Appli
Sequence 16, Appl
Sequence 1, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 868, App
Sequence 1016, Ap
Sequence 853, App
Sequence 820, App
Sequence 385, App
Sequence 1299, Ap
Sequence 1426, Ap

ALIGNMENTS

late

4

QY 148 CysSerSerAsnGlyArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGly 167
 Db 68 TGCTCCAGTAATGGCGGTAGAGGCGCGAGCAGCAAGAAATCGGCGTTTACTGGGGCCAGGC 9
 QY 168 HisPro 169
 Db 8 CATCCT 3

RESULT 3
 US-10-040-862-9423/c
 ; Sequence 9423, Application US/10040862
 ; Publication No. US20030078396A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gaiger, Alexander
 ; APPLICANT: Algate, Paul A.
 ; APPLICANT: Mannion, Jane
 ; APPLICANT: Retter, Marc
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
 ; TITLE OF INVENTION: Hematological Malignancies
 ; FILE REFERENCE: 014058-013520US
 ; CURRENT APPLICATION NUMBER: US/10/040,862
 ; CURRENT FILING DATE: 2001-11-06
 ; PRIOR APPLICATION NUMBER: US 60/186,126
 ; PRIOR FILING DATE: 2000-03-01
 ; PRIOR APPLICATION NUMBER: US 60/190,479
 ; PRIOR FILING DATE: 2000-03-17
 ; PRIOR APPLICATION NUMBER: US 60/200,545
 ; PRIOR FILING DATE: 2000-04-27
 ; PRIOR APPLICATION NUMBER: US 60/200,303
 ; PRIOR FILING DATE: 2000-04-28
 ; PRIOR APPLICATION NUMBER: US 60/200,779
 ; PRIOR FILING DATE: 2000-04-28
 ; PRIOR APPLICATION NUMBER: US 60/200,999
 ; PRIOR FILING DATE: 2000-05-01
 ; PRIOR APPLICATION NUMBER: US 60/202,084
 ; PRIOR FILING DATE: 2000-05-04
 ; PRIOR APPLICATION NUMBER: US 60/206,201
 ; PRIOR FILING DATE: 2000-05-22
 ; PRIOR APPLICATION NUMBER: US 60/218,950
 ; PRIOR FILING DATE: 2000-07-14
 ; PRIOR APPLICATION NUMBER: US 60/222,903
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: US 60/223,416
 ; PRIOR FILING DATE: 2000-08-04
 ; PRIOR APPLICATION NUMBER: US 60/223,378
 ; PRIOR FILING DATE: 2000-08-07
 ; PRIOR APPLICATION NUMBER: US 09/796,692
 ; PRIOR FILING DATE: 2001-03-01
 ; NUMBER OF SEQ ID NOS: 10467
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 9423
 ; LENGTH: 310
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-040-862-9423

Alignment Scores:
 Pred. No.: 4,13e-56 Length: 310
 Score: 549.00 Matches: 98
 Percent Similarity: 98.04% Conservative: 2
 Best Local Similarity: 96.08% Mismatches: 2
 Query Match: 35.51% Indels: 0
 DB: 15 Gaps: 0

US-10-054-313-1 (1-286) x US-10-040-862-9423 (1-310)

QY 68 TrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis 87
 Db 308 TGGGCGCTTTGTCAGGAATCTCAGAGCCGCGAAGTTTCAGAGGCGCATGAAATCAACAT 249
 QY 88 GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGly 107

Db 248 GGCAAGAATCGGAGCGAAAGCCAGCAAGCGACTCCGTGAGCCACTGGATGGAGATGGA 189
 QY 108 HisGluSerAsnGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127
 Db 188 CATGAAGCGCAGAGCGGTATGCAAGACACATGAAGCCGCGTGGAGCGCGCCCTCCA 129
 QY 128 ValSerArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCys 147
 Db 128 GTTAGCAGAGACACGTTTCTTACATGGGAGACTTCGTGCTGTCTACACTGATGCTGC 69
 QY 148 CysSerSerAsnGlyArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGly 167
 Db 68 TGCTCCAGTAATGGCGGTAGAGGCGCGAGCAGCAAGAAATCGGCGTTTACTGGGGCCAGGC 9
 QY 168 HisPro 169
 Db 8 CATCCT 3

RESULT 4
 US-09-796-692-7011/c
 ; Sequence 7011, Application US/09796692
 ; Publication No. US20020198362A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gaiger, Alexander
 ; APPLICANT: Algate, Paul A.
 ; APPLICANT: Mannion, Jane
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAP
 ; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
 ; FILE REFERENCE: 2077.001200
 ; CURRENT APPLICATION NUMBER: US/09/796,692
 ; CURRENT FILING DATE: 2001-03-01
 ; PRIOR APPLICATION NUMBER: 60/186,126
 ; PRIOR FILING DATE: 2000-03-01
 ; PRIOR APPLICATION NUMBER: 60/190,479
 ; PRIOR FILING DATE: 2000-03-17
 ; PRIOR APPLICATION NUMBER: 60/200,545
 ; PRIOR FILING DATE: 2000-04-27
 ; PRIOR APPLICATION NUMBER: 60/200,303
 ; PRIOR FILING DATE: 2000-04-28
 ; PRIOR APPLICATION NUMBER: 60/200,779
 ; PRIOR FILING DATE: 2000-04-28
 ; PRIOR APPLICATION NUMBER: 60/200,999
 ; PRIOR FILING DATE: 2000-05-01
 ; PRIOR APPLICATION NUMBER: 60/202,084
 ; PRIOR FILING DATE: 2000-05-04
 ; PRIOR APPLICATION NUMBER: 60/206,201
 ; PRIOR FILING DATE: 2000-05-22
 ; PRIOR APPLICATION NUMBER: 60/218,950
 ; PRIOR FILING DATE: 2000-07-14
 ; PRIOR APPLICATION NUMBER: 60/222,903
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: 60/223,416
 ; PRIOR FILING DATE: 2000-08-04
 ; PRIOR APPLICATION NUMBER: 60/223,378
 ; PRIOR FILING DATE: 2000-08-07
 ; NUMBER OF SEQ ID NOS: 9597
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 7011
 ; LENGTH: 310
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-796-692-7011

Alignment Scores:
 Pred. No.: 3,69e-55 Length: 310
 Score: 541.00 Matches: 97
 Percent Similarity: 97.06% Conservative: 2
 Best Local Similarity: 95.10% Mismatches: 3
 Query Match: 34.99% Indels: 0
 DB: 10 Gaps: 0

US-10-054-313-1 (1-286) x US-09-796-692-7011 (1-310)

Qy 68 TtpAlaPheValAcLysSerProGluValSerGluGluHis 87
Db 308 TGGGCTTTGTGAGGAATCTTCAAGCCCGGAAGTTTCAGAGGCGCATGAAATCAACAT 249
Qy 88 GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGly 107
Db 248 GGACAAGAATCGAGCGGAAAGCCAGCAAGCGACTCCCGTGAGCCACTGGATGGAGATGGA 189
Qy 108 HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127
Db 188 CATGAAGCGCGAGCGCGTATGCAAGCACATGAATCGACGCTGGAGCCCGCCCTCCA 129
Qy 128 ValSerArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCys 147
Db 128 GTTACGAGAGACACGTTTCTTACATGGAGACTTCGTCGTCTACACTGATGGCTGC 69
Qy 148 CysSerSerAsnGlyArgArgLysProArgAlaGlyLleGlyValTyrTyrGlyProGly 167
Db 68 TGCTCCAGTAATGGCGGTAGAGGCGCGGAGCAGGAATCGGCGTTTACTGGGGGCGCGGC 9
Qy 168 HisPro 169
Db 8 CATCCT 3
RESULT 5
US-10-040-862-7011/c
; Sequence 7011, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Hematological Malignancies
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7011
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-040-862-7011

Alignment Scores: 3.69e-55 Length: 310
Pred. No.: 541-00 Matches: 97
Score: 97.06% Conservatives: 2
Best Local Similarity: 95.10% Mismatches: 3
Query Match: 34.99% Indels: 0
DB: 15 Gaps: 0
US-10-054-313-1 (1-286) x US-10-040-862-7011 (1-310)
Qy 68 TtpAlaPheValAcLysSerProGluValSerGluGluHis 87
Db 308 TGGGCTTTGTGAGGAATCTTCAAGCCCGGAAGTTTCAGAGGCGCATGAAATCAACAT 249
Qy 88 GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGly 107
Db 248 GGACAAGAATCGAGCGGAAAGCCAGCAAGCGACTCCCGTGAGCCACTGGATGGAGATGGA 189
Qy 108 HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127
Db 188 CATGAAGCGCGAGCGCGTATGCAAGCACATGAATCGACGCTGGAGCCCGCCCTCCA 129
Qy 128 ValSerArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCys 147
Db 128 GTTACGAGAGACACGTTTCTTACATGGAGACTTCGTCGTCTACACTGATGGCTGC 69
Qy 148 CysSerSerAsnGlyArgArgLysProArgAlaGlyLleGlyValTyrTyrGlyProGly 167
Db 68 TGCTCCAGTAATGGCGGTAGAGGCGCGGAGCAGGAATCGGCGTTTACTGGGGGCGCGGC 9
Qy 168 HisPro 169
Db 8 CATCCT 3
RESULT 6
US-09-918-995-29506
; Sequence 29506, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29506
; LENGTH: 436
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(436)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-29506
Alignment Scores: 2.77e-45 Length: 436
Pred. No.: 468-00 Matches: 86
Score: 100.00% Conservatives: 3
Best Local Similarity: 96.63% Mismatches: 0
Query Match: 30.27% Indels: 0
DB: 11 Gaps: 0
US-10-054-313-1 (1-286) x US-09-918-995-29506 (1-436)
Qy 198 LysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGly 217
Db 32 CAGACTCAAAACATCAATAACTGGTTCTGTATACAGACAGTATGTTTACTTTAAATGGT 91

Qy 218 IleThrAsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGlu 237
Db 92 ATTTCTAACTGGGTTCAAGGTTGGAAGAAAAATGGTGGAGACAAGTGCAGGGAAGAG 151
Qy 238 ValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGln 257
Db 152 GTGATCAACAAGAGAGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAG 211
Qy 258 TrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAla 277
Db 212 TGGATGCATGTTCTCTGGTCATTCGGGATTTATAGCAATGAAGAAGCTGCAGATTAGCC 271
Qy 278 ArgGluGlyAlaLysGlnSerGluAsp 286
Db 272 AGAGAAGGAGCTAAACAATCGGAAGAC 298

RESULT 7

US-09-814-353-19582
; Sequence 19582, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-0065
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19582
; LENGTH: 1279
; TYPE: DNA
; ORGANISM: Homo sapiens

FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1262, 1263, 1264, 1265, 1266, 1267, 1268, 1269, 1270, 1271,
; LOCATION: 1272, 1273, 1274, 1275, 1276, 1277, 1278, 1279
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-19582

Alignment Scores:
Pred. No.: 1.638-43 Length: 1279
Score: 450.00 Matches: 110
Percent Similarity: 35.83% Conservative: 0
Best Local Similarity: 35.83% Mismatches: 1
Query Match: 29.11% Indels: 197
DB: 13 Gaps: 1

US-10-054-313-1 (1-286) x US-09-814-353-19582 (1-1279)

Qy 171 AsnValGlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAla 190
Db 21 AATGTAGGCATTAGACTTCTCGGGCGGAGACAAACCAAGAGCGAAATTCAGCGCC 80
Qy 191 CysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyThrAsp 210
Db 81 TCGAAGCCATTGAAACAGCAAGAACTCAAAACATCAATAAAGTGGTCTGTATACAGAC 140

Qy 211 SerMetPheThrIleAsnGlyIle----- 218
Db 141 AGTATGTTTACGATAAATGGTAA-GCTTTACATTTGATTTCTGTTTCTTCCAGTAAC 199
Qy 218 ----- 218
Db 200 TGTGAAGGAAATTTGGTAGGAGGTGTTGTAACAGGCGAGGCCAAATCGGAACCGGGGG 259
Qy 218 ----- 218
Db 260 ATGACATTGGTTGTGTCAGGTACCGAGCAAGAGTGAAGATTTTGGAGTCTCCCTTCTGCT 319
Qy 218 ----- 218
Db 320 GCTCTGATGTTTTCCACATGCTTATTTCTTCCAGGCACTGGAGATCGATCAGAAAGTG 379
Qy 218 ----- 218
Db 380 GAAGTGGCTCTTACTTCTAGTCTGTGTGTATAGTCACTTAAGATGCGGTGTTGACTG 439
Qy 218 ----- 218
Db 440 CTTCTTTGGAAATGCCCTGAATAGGAGCATGTAGGGGATGCTTACGAGGCTGGGGAAG 499
Qy 218 ----- 218
Db 500 CTCTCGGAGAAATGACATCTTAAGCTGACAGCTGTAGTAGAATACAAGTTAGGGGAAA 559
Qy 218 ----- 218
Db 560 GCGAAGAGAGAGGGTATTTTCAGGCTGTGCCAAGGCCAAAGCAACTAAGCGCAGCTAAG 619
Qy 218 ----- 218
Db 620 GAAGTGAAGAGTCAAGGATGGTGACTCAGATACAGTGGGATCAATTATAGAGGTAAG 679
Qy 218 ----- 218
Db 680 ACTAAGGAGGTGAAAAAGGCGAGCACACTCAAGATTGACTCACCAGTGTGTCATTTT 739
Qy 219 -----ThrAsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGly 235
Db 740 GTAGGTATTAACCTAACTGGGTTCAAGGTTGGAAGAAAAATGGTGGGAAGACAAGTGCAGG 799
Qy 236 LysGlu-ValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAs 255
Db 800 AAAGAGGGTGATCAACAAGAGGAGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGA 859
Qy 255 pIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGlnAlaAsp 275
Db 860 CATTCAGTGGATGCATGTTCTCTGGTCAATTCGGGATTTATAGGCAATGAAGAAGCTGACAG 919
Qy 275 gLeuAlaArgGluGlyAla 281
Db 920 ATTAGCCAGAGAGGAGCA 938

RESULT 8

US-09-960-352-6980
; Sequence 6980, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Ningbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21 (10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 6980
; LENGTH: 473

```
/ TYPE: DNA
/ ORGANISM: Bos taurus
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (404)
/ OTHER INFORMATION: unsure at all n locations
/ OTHER INFORMATION: Clone ID: 30-LIB3058-022-Q1-K1-H5
US-09-960-352-6980

Alignment Scores:
Pred. No.: 3,4e-42 Length: 473
Score: 434.00 Matches: 79
Percent Similarity: 88.78% Conservative: 8
Best Local Similarity: 80.61% Mismatches: 11
Query Match: 28.07% Indels: 0
DB: 10 Gaps: 0

US-10-054-313-1 (1-286) x US-09-960-352-6980 (1-473)
Qy 189 AlaalaCysysAlaileGluGlnAlaLysThrGlnAsnleAsnLysLeuValLeuTyr 208
Db 79 GCAGCCTGCAAGCCATAGACGAGCTTAGGCTCAGGACATCACTAAGCTGTTCTCTAC 138
Qy 209 ThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLysAsn 228
Db 139 ACAGACAGCATGTTTACCATCAATGGCATCCACCCTGGTGAAGGCTGGAAGCAGAT 198
Qy 229 GlyTrpLysThrSerAlaGlyIysGluValIleAsnLysGluAspPheValAlaLeuGlu 248
Db 199 GCGTGGAGGACCAACGACCGGAGGAGGTGACCAACAGAGGACTTCGGGAGCTGGAG 258
Qy 249 ArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIle 268
Db 259 CGGCTGGCGGGGCGATGACATTCAGTGGATGCACGTTCTTGGCCATTCCGGGTTTAA 318
Qy 269 GlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAlaLysGlnSerGluAsp 286
Db 319 GGCAATGAAAGGCGGACAGACTATCGAGAGAGGTGCAAAACACTCCCGCAGAC 372

RESULT 9
US-09-960-352-3900
/ Sequence 3900, Application US/09960352
/ Patent No. US20020137139A1
/ GENERAL INFORMATION:
/ APPLICANT: Warren, Wesley C.
/ APPLICANT: Tao, Nengbing
/ APPLICANT: Byatt, John C.
/ APPLICANT: Mathialagan, Nagappan
/ TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
/ FILE REFERENCE: 1511.006/37-21(10298)C
/ CURRENT APPLICATION NUMBER: US/09/960,352
/ CURRENT FILING DATE: 2001-09-24
/ NUMBER OF SEQ ID NOS: 15112
/ SEQ ID NO 3900
/ LENGTH: 407
/ TYPE: DNA
/ ORGANISM: Bos taurus
/ OTHER INFORMATION: Clone ID: 17-LIB3058-058-Q1-K1-E1
US-09-960-352-3900

Alignment Scores:
Pred. No.: 4.01e-38 Length: 407
Score: 399.00 Matches: 81
Percent Similarity: 75.81% Conservative: 13
Best Local Similarity: 65.32% Mismatches: 24
Query Match: 25.81% Indels: 6
DB: 10 Gaps: 2

US-10-054-313-1 (1-286) x US-09-960-352-3900 (1-407)
Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20

Db 52 ATGACCCGGCTGCTGGCGGTGGTGCCACAGAGTGGCTTTGGCCCTTTGGGATGC----- 105
Qy 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgGlyArgGlyThrGlyValPhe 40
Db 106 -----CGCAGTCTTGCCATGTTCTATGCGTGAGGAGGGCGCAAGGCGGGGTCTTC 159
Qy 41 LeuTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db 160 CTGACCTGGGATGAATCGAGACAGCAGTGGACCGCTTCTCGAGCCAGATTAAAG 219
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 220 TTCGCCACAGAGGAGGAGGCTGGCGCTTTGTGAGAGGTCTGAAAGCCCGCAGTATCA 279
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGlnSerGluAlaLysProGlyLysArgLeuArg 100
Db 280 GAAGGCAAAACTAAACACGCTAGAAAGAACCCACAGCGAAAGCGGACCAAGCGACCTCCGT 339
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 340 GAGCCCTTGGATGAAGAT-----GAAGATGCAAAAGCGGTGTGCGAAGCACGTGCGGGAG 393
Qy 121 SerValGluPro 124
Db 394 AGCGCAGAGCCA 405

RESULT 10
US-09-910-943-431
/ Sequence 431, Application US/09910943
/ Patent No. US20020081610A1
/ GENERAL INFORMATION:
/ APPLICANT: Hemmati-Briavanlou, Ali
/ APPLICANT: Altman, Curtis
/ TITLE OF INVENTION: Assays and Materials for Embryonic Gene Expression
/ FILE REFERENCE: 7529/IG148US1
/ CURRENT APPLICATION NUMBER: US/09/910,943
/ CURRENT FILING DATE: 2001-07-23
/ NUMBER OF SEQ ID NOS: 742
/ SOFTWARE: Patent in version 3.1
/ SEQ ID NO 431
/ LENGTH: 764
/ TYPE: DNA
/ ORGANISM: Xenopus laevis
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)..(764)
/ OTHER INFORMATION: n may be a or g or c or t/u
US-09-910-943-431

Alignment Scores:
Pred. No.: 1.36e-33 Length: 764
Score: 364.00 Matches: 83
Percent Similarity: 58.38% Conservative: 18
Best Local Similarity: 47.98% Mismatches: 57
Query Match: 23.54% Indels: 15
DB: 9 Gaps: 5

US-10-054-313-1 (1-286) x US-09-910-943-431 (1-764)
Qy 27 MetPheTrpAlaValArgArgGlyArgLysThrGlyValPheLeuThrTrpAsnGluCys 46
Db 249 ATGTTTACGCTGTAAGGACTGGCGCTGAAAGCGCTGAGTCTACATACGTTGGGATGATGT 308
Qy 47 ArgAlaGlnValAspArgPheProAlaAlaArgPheLysLysPheAlaThrGluAspGlu 66
Db 309 AAAGAGCAAGTGGATCGATTTCCTTTAGCAAGGTTCAGAGAAGTTTGCTTCAGAGGAAGAT 368
Qy 67 AlaTrpAlaPheValArg-----LysSerAlaSerProGluValSerGluGlyHisGlu 84
Db 369 GCCTGGGAATTTGTGAGGAACACTCAGGAATCATCATCGAAAGGTCTTACTAGTGTGAA 428
Qy 85 AsnGlnHisGlyGlnGluSerGluAlaLys-----ProGly 96
Db 85 AsnGlnHisGlyGlnGluSerGluAlaLys-----ProGly 96
```

```

Db 429 ACAAGGAGCCTCTACACAGCTACAAAAGCTGCAGGACTGCATATATGTCATACCTCAG 498
Qy 97 LysArgIeuArgGluPro---LysAspGlyAspGlyHisGluSerAlaGlnProTyrAla 115
Db 489 TCCAGAGAAAGACCACTACTACAGAGCTCAAGCACTGAGAAGAGCATCTCACCTTAAA 548
Qy 116 LysHisMetLysProSerValGluProAlaProValSerArgAspThrPheSerTyr 135
Db 549 AGAAGCAAGCTCATGTATCATCTGATTCATCATCTCGTCACATATGAACCTTTTACCTAC 608
Qy 136 MetGlyAppheValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLys 155
Db 609 ATGGAGACGCTGCTGTGTATACACTGATCGCTGTAGCGGAATGGCGGGTAAAA 668
Qy 156 ProArgAlaGlyIleGlyValTyrTrp-GlyProGlyHisProLeuAsnValGlyIleAr 175
Db 669 GCACAGCTGTATAGGTGTATATCTGGGGGCAAGGCGCTTCTCTNAACTTGCGAGAAA 728
Qy 175 gLeuProGlyArgGlnThrAsn-----GlnArgAla 185
Db 729 GCTT---GGAAGGAGGCAAACTTAACCCAGCGGGCT 762

RESULT 11
US-09-814-353-13698
; Sequence 13698, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13698
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-13698

Alignment Scores:
Pred. No.: 4,178-19 Length: 499
Score: 240.00 Matches: 47
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.52% Indels: 0
DB: 13 Gaps: 0

US-10-054-313-1 (1-286) x US-09-814-353-13698 (1-499)
; Sequence 7313, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B

```

```

Qy 211 SerMetPheThrIleAsnGly 217
Db 140 AGTATGTTTACGATAAATGGT 160

RESULT 12
US-09-814-353-943
; Sequence 943, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 943
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-943

Alignment Scores:
Pred. No.: 4,19e-19 Length: 501
Score: 240.00 Matches: 47
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.52% Indels: 0
DB: 13 Gaps: 0

US-10-054-313-1 (1-286) x US-09-814-353-943 (1-501)
Qy 171 AsnValGlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAla 190
Db 19 AATGTAGGCATTAGACTTCTCTGGCGGAGACAAACCAAGAGCGGAAATTCATGCAGCC 78
Qy 191 CysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAsp 210
Db 79 TGCAGAGCCATTGACCAAGCAAGACTCAAAACATCAATAAACTGTTCTGTATACAGAC 138

Qy 211 SerMetPheThrIleAsnGly 217
Db 139 AGTATGTTTACGATAAATGGT 159

RESULT 13
US-09-814-353-7313
; Sequence 7313, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B

```

```

; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-08-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7313
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-7313

```

```

Alignment Scores:
Pred. No.: 4,19e-19 Length: 501
Score: 240.00 Matches: 47
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15,52% Indels: 0
DB: 13 Gaps: 0

```

```
US-10-054-313-1 (1-286) x US-09-814-353-7313 (1-501)
```

```

Qy 171 AsnValcIleArgLeuProGlyAArgGlnThrAsnGlnArgAlaGluLeuHisAlaAla 190
Db 19 ATGTAGGCGATAGACTTCTGGCGGCGAGACAAACCAAGCGGAATTCATCGAGCC 78
Qy 191 CysLysAlaIleGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAsp 210
Db 79 TGCAGAGCCATTGAACAAGCAAGACTCAAAACATCAATAAACTGGTTCTGTATACAGAC 138
Qy 211 SerMetPheThrIleAsnGly 217
Db 139 AGTATGTTTACGATAAATGGT 159

```

```
RESULT 14
```

```
US-10-260-877-71
```

```

; Sequence 71, Application US/10260877
; Publication No. US20030021813A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Chovan, Linda E.
; APPLICANT: Hessler, Paul E.
; APPLICANT: Reich, Karl A.
; TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME
; TITLE OF INVENTION: SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: 'ESSENTIAL GENES'
; FILE REFERENCE: 6565 US.P1
; CURRENT APPLICATION NUMBER: US/10/260,877
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US/09/649,145
; PRIOR FILING DATE: 2000-08-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 462
; TYPE: DNA
; ORGANISM: H. influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(462)
; OTHER INFORMATION: HI-0138
; FEATURE:
; NAME/KEY: misc_feature

```

```

; LOCATION: (6)...(6)
; OTHER INFORMATION: k = g or t/u at position 6
US-10-260-877-71

```

```

Alignment Scores:
Pred. No.: 9,79e-19 Length: 462
Score: 236.50 Matches: 54
Percent Similarity: 49.68% Conservative: 23
Best Local Similarity: 34.84% Mismatches: 49
Query Match: 15.30% Indels: 29
DB: 15 Gaps: 4

```

```
US-10-054-313-1 (1-286) x US-10-260-877-71 (1-462)
```

```

Qy 140 ValValValTyrThrAspGlyCysSerSerAsnGlyArgGlySerProArgAlaGly 159
Db 13 ATTGAAATTTTACTGATGATCTTCTGTTAGGTAAT-----CCAGGGCGGGC 60
Qy 160 IleGlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArg 179
Db 61 -----GGAATTGGTCCCGTATTGCGTTATAAA 87
Qy 180 Gln-----ThrAsnGlnArgAlaGluLeu 187
Db 88 CAACATGAAAAACACTCTCCAAAGCGTATTTCCAAACCCCAATTAATCGAATGGGAATTA 147
Qy 188 HisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeu 207
Db 148 CGCGCTGTCATTGAGCATTAAATACATTAAAGAACCTTGCTTG-----ATCACGCTT 201
Qy 208 TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTyrLysLys 227
Db 202 TATAGTCATAGCAATATATGAAAAATGGCATAACCAAAATGGATCTTTAACTGGAAAAAA 261
Qy 228 AsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeu 247
Db 262 AATAATTGGAAAGCAAGTCTCGAAGACCTGTAAAAAACCAAGATTTATGGATAGCTTA 321
Qy 248 GluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe 267
Db 322 GATGATCCATCCAAACGTCATAAAATTAATTGGCAATGGTAAAGGCCATGCTGGACAC 381
Qy 268 IleGlyAsnGluAlaAspArgLeuAlaArgGluGlyAlaLys 282
Db 382 AGAGAAATGAATTTCCGATGATAGCAAAAGGGGCGAGAA 426

```

```
RESULT 15
```

```
US-10-329-960-1
```

```

; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c

```

```
FEATURE:
NAME/KEY: misc feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (29238)..(29238)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
```

```
NAME/KEY: misc feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
```

Search completed: December 17, 2003, 22:32:49
Job time : 955 secs

```
LOCATION: (142750)..(142750)
FEATURE:
OTHER INFORMATION: n equals a, t, g or c
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152530)..(152530)

Alignment Scores:
Pred. No.: 6.98e-14 Length: 1830:21
Score: 236.50 Matches: 54
Percent Similarity: 49.68% Conservative: 23
Best Local Similarity: 34.84% Mismatches: 49
Query Match: 15.30% Indels: 29
DB: 15 Gaps: 4

US-10-054-313-1 (1-286) x US-10-329-960-1 (1-1830121)
Qy 140 ValValValValThrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGly 159
Db 152742 ATTGAATTTTACTGATGATCTTCTAGTAAAT-----CCAGGGGGGGGC 152789
Qy 160 IleGlyValTrpTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArg 179
Db 152790 -----GGAATTGGTCCGCTATTGCGTTATAAA 152816
Qy 180 Gln-----ThrAsnGlnArgAlaGluIle 187
Db 152817 CAATGAAAAAACACTCTCAAGAGCTATTTCACAAACCCCAATTAATCGAATGGAAATTA 152876
Qy 188 HisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeu 207
Db 152877 CGCGCTGCTCAATGAAGCATTAAATACATTAAAGAACCTTGCTTG-----ATCAGCCTT 152930
Qy 208 TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLys 227
Db 152931 TATAGTCATGCCAATATATCAAAAAATGCGATAACCAAAATGGATCTTTAACTGAAAAAAA 152990
Qy 228 AsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeu 247
Db 152991 AATAATTTGGAAGCAAGCTCTGGAAGCCCTGTAAGAACCAAGATTTATGGATAGCCTTA 153050
Qy 248 GluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe 267
Db 153051 GATGAATCCATCCCAACCTCAATAAATTAATGGCAATGGGTAAAGGCCCATGCTGGACAC 153110
Qy 268 IleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAlaLys 282
Db 153111 AGAGAAATGAAATTTGCGATGAATTAGCAAAAAAGGGGCAGAA 153155
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 17, 2003, 19:23:44 : Search time 285 Seconds

(without alignments)
2002.513 Million cell updates/sec

Title: US-10-054-313-1

Perfect score: 1546

Sequence: 1 MSWLLFLAHRVALAALPCRR.....FIGNEEDRLAREGAKQSED 286

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3357611 seqs, 997755592 residues

Total number of hits satisfying chosen parameters: 6715222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10054313/runat.17122003.150746.24307/app.query.fasta.1.455
-DS=Pending Patents NA New -FMFT=fastap -SUFFIX=fnpn -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFM=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054313 @CEN 1.1.282 @runat.17122003.150746.24307 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Pending Patents NA New:
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|---------|----|---------------------------------------|
| 1 | 1536 | 98.7 | 1168 | 7 | US-60-490-890-1779 Sequence 1779, Ap |
| 2 | 549 | 35.5 | 310 | 5 | US-09-796-692A-9423 Sequence 9423, Ap |
| 3 | 541 | 35.0 | 310 | 5 | US-09-796-692A-7011 Sequence 7011, Ap |
| 4 | 419 | 27.1 | 629 | 7 | US-60-507-481-9585 Sequence 9585, Ap |
| 5 | 272 | 17.6 | 119211 | 6 | US-10-649-136-40 Sequence 40, Appl |
| 6 | 272 | 17.6 | 119211 | 6 | US-10-672-787-40 Sequence 40, Appl |
| 7 | 229.5 | 14.8 | 36471 | 5 | US-09-806-866A-1 Sequence 1, Appl |
| 8 | 229.5 | 14.8 | 2242716 | 5 | US-09-806-866A-1068 Sequence 1068, Ap |
| 9 | 212.5 | 13.7 | 1733 | 5 | US-09-897-516A-658 Sequence 658, App |
| 10 | 212.5 | 13.7 | 1733 | 5 | US-09-897-516A-659 Sequence 659, App |
| 11 | 212.5 | 13.7 | 1733 | 5 | US-09-897-516A-660 Sequence 660, App |

| | | | | | |
|----|-------|------|----------|---|--|
| 12 | 208.5 | 13.5 | 1592 | 6 | US-10-684-141-41 Sequence 41, Appl |
| 13 | 194.5 | 12.6 | 2731748 | 6 | US-10-297-465B-1 Sequence 1, Appl |
| 14 | 171.5 | 11.1 | 1545 | 6 | US-10-425-114A-34337 Sequence 34337, A |
| 15 | 162.5 | 10.5 | 644 | 6 | US-10-653-047-5466 Sequence 5466, Ap |
| 16 | 156 | 10.1 | 8889 | 6 | US-10-677-558-4 Sequence 4, Appl |
| 17 | 156 | 10.1 | 11364 | 6 | US-10-677-558-11 Sequence 11, Appl |
| 18 | 156 | 10.1 | 11394 | 6 | US-10-677-558-5 Sequence 5, Appl |
| 19 | 139.5 | 9.0 | 270 | 5 | US-09-674-546A-183 Sequence 183, App |
| 20 | 138 | 8.9 | 270 | 5 | US-09-674-546A-181 Sequence 181, App |
| 21 | 138 | 8.9 | 270 | 5 | US-09-674-546A-185 Sequence 185, App |
| 22 | 137 | 8.9 | 2151 | 1 | PCT-US03-38802-1 Sequence 1, Appl |
| 23 | 137 | 8.9 | 2151 | 6 | US-10-661-819-1 Sequence 1, Appl |
| 24 | 134.5 | 8.7 | 9464 | 6 | US-10-673-023-1 Sequence 1, Appl |
| 25 | 134.5 | 8.7 | 9464 | 6 | US-10-673-008-1 Sequence 1, Appl |
| 26 | 134.5 | 8.7 | 9464 | 6 | US-10-702-755-1 Sequence 1, Appl |
| 27 | 134.5 | 8.7 | 9464 | 6 | US-10-672-489-1 Sequence 1, Appl |
| 28 | 125 | 8.1 | 8059021 | 5 | US-09-947-914-53 Sequence 53, Appl |
| 29 | 124 | 8.0 | 8003 | 6 | US-10-276-869A-38 Sequence 38, Appl |
| 30 | 121 | 7.8 | 111190 | 7 | US-60-500-315-11870 Sequence 11870, A |
| 31 | 121 | 7.8 | 184000 | 7 | US-60-495-114-16699 Sequence 16699, A |
| 32 | 121 | 7.8 | 1748349 | 5 | US-09-947-914-48 Sequence 48, Appl |
| 33 | 116 | 7.5 | 40260 | 7 | US-60-500-315-11639 Sequence 11639, A |
| 34 | 115.5 | 7.5 | 16360 | 1 | PCT-US03-20325-1 Sequence 1, Appl |
| 35 | 115.5 | 7.5 | 17207 | 1 | PCT-US03-20325-8 Sequence 8, Appl |
| 36 | 113.5 | 7.3 | 469219 | 7 | US-60-495-114-16704 Sequence 16704, A |
| 37 | 113 | 7.3 | 21744 | 7 | US-60-500-315-11634 Sequence 11634, A |
| 38 | 112.5 | 7.3 | 26021 | 7 | US-60-495-114-16436 Sequence 16436, A |
| 39 | 112 | 7.2 | 141158 | 6 | US-10-719-993-6837 Sequence 6837, Ap |
| 40 | 111.5 | 7.2 | 244289 | 7 | US-60-495-114-16866 Sequence 16866, A |
| 41 | 111 | 7.2 | 1203 | 1 | PCT-US03-33947-307 Sequence 307, App |
| 42 | 111 | 7.2 | 1203 | 1 | PCT-US03-33947-1875 Sequence 1875, App |
| 43 | 111 | 7.2 | 25821 | 7 | US-60-487-610-20089 Sequence 20089, A |
| 44 | 111 | 7.2 | 25821 | 7 | US-60-485-450-12348 Sequence 12348, A |
| 45 | 109 | 7.1 | 13831263 | 5 | US-09-947-914-41 Sequence 41, Appl |

ALIGNMENTS

RESULT 1
US-60-490-890-1779
; Sequence 1779, Application US/60490890
; GENERAL INFORMATION:
; APPLICANT: Li, Martha
; APPLICANT: Rupnow, Brent A.
; APPLICANT: Webster, Kevin R.
; APPLICANT: Jackson, Donald
; APPLICANT: Wong, Tai W.
; TITLE OF INVENTION: BIOMARKERS OF CYCLIN-DEPENDENT KINASE MODULATION
; FILE REFERENCE: D0310 PSP
; CURRENT APPLICATION NUMBER: 294607490, 890
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 2779
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 1779
; LENGTH: 1168
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-490-890-1779

Alignment Scores:
Pred. No.: 3.66e-172 Length: 1168
Score: 1536.00 Matches: 282
Percent Similarity: 99.30% Conservative: 2
Best Local Similarity: 98.60% Mismatches: 2
Query Match: 98.71% Indels: 0
DB: 7 Gaps: 0

US-10-054-313-1 (1-286) x US-60-490-890-1779 (1-1168)

Oy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20

Db 82 ATGAGCTGGCTTCGTCTTCCTGGCCACAGAGTCGCCTTGGCCGCTTGCCTCCGCCGC 141

```
Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValaArgGlyArgGlyThrGlyValPhe 40
Db 142 GGCCTTCGGGGTTTCGGGATGTTCTATCCGTGAGAGGGCCGCGAAGCCGGGGTCTTT 201
Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValaAspArgPheProAlaAlaArgPheLys 60
Db 202 CTGACCTGGATGAGTCAGAGCACAGGTGGACCGGTTCTCTGCTGCCAGATTTAAGAAG 261
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 262 TTGGCCACAGAGATGAGCGCTTGGCCCTTGTGAGGAATCTGCAAGCCCGGAGATTTC 321
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 322 GAAGGCGATGAATCAATCAATGACATGACAGAAATCGGAGCGGAAAGCCAGCAAGCGCTCGT 381
Qy 101 GluProLeuAspGlyAspGlyHisGlnGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 382 GAGCCACTGGATGGAGATGGACATGAAGCGCAGAGCCGTATGCAAGCCAGCATGAAGCGG 441
Qy 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 442 AGCGTGGAGCCGCGCTCCAGTTAGCAGACACAGTTTCTACATCGGAGACTTCGTC 501
Qy 141 ValValThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
Db 502 GTCGCTACACTGATGGCTGCTCCAGTAATGGCGGTAGAGAGCCGCGAGCAAGCAATC 561
Qy 161 GlyValThrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db 562 GACGTTTACTGGGGCCAGGCCATCTTTAAATGTAGGCAATTAGACTTCTCTGGGGCGGAG 621
Qy 181 ThrAsnGlnArgAlaGluIleHisAlaCysLysAlaIleGluGlnAlaLysThrGln 200
Db 622 ACAACCAAGAGCGGAAATTCATGCGCTGCAAGCCATTGAAACAAAGCAAGCACTCAA 681
Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 682 AACATCAATAACTGGTCTCTGATACAGACATATGTTTACGATAAATGGTATATACTAAC 741
Qy 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
Db 742 TGGGTTCAAGGTTGGAAGAAAATGGTGGAGAGCAAGTGCAGGCAAGAGGTGATCAAC 801
Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db 802 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCAGGGGATGGACATTCAGTGGATGCAT 861
Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db 862 GTTCCTGGTCATTCGGGATTTATAGCAATGAAGAGCTGACAGATTAGCCAGAGAGGA 921
Qy 281 AlaLysGlnSerGluAsp 286
Db 922 GCTAAACAAATCGGAAGAC 939
```

RESULT 2

```
US-09-796-692A-9423/c
; Sequence 9423, Application US/09796692A
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013512US
; CURRENT APPLICATION NUMBER: US/09/796,692A
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
```

```
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9423
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692A-9423

Alignment Scores:
Pred. No.: 4,78e-56 Length: 310
Score: 549.00 Matches: 98
Percent Similarity: 98.04% Conservative: 2
Best Local Similarity: 96.08% Mismatches: 2
Query Match: 35.51% Indels: 0
DB: Gaps: 0

US-10-054-313-1 (1-286) x US-09-796-692A-9423 (1-310)
Qy 68 TrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis 87
Db 308 TGGGCTTTCTCAGGAATCTGCAAGCCCGGAAGTTTCAGAAGGCGATGAAATCAACAT 249
Qy 88 GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgLysProLeuAspGlyAspGly 107
Db 248 GGAACAAGATCGGAGGCGAAAGCCAGCAAGCGACTCCGTGAGCCACTGGATGGAGATGA 189
Qy 108 HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127
Db 188 CATGAAGCCAGAGCCGCTATGCAAGCACATGAAGCCGCGCTGGAGCGCGCCCTCCA 129
Qy 128 ValSerArgAspThrPheSerTyrMetGlyAspPheValValValThrAspGlyCys 147
Db 128 GTTAGCAGACACAGTTTCTCTACATGGGAGACTTCGTCGCTCTACACTGATGGCTGC 69
Qy 148 CysSerSerAsnGlyArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGly 167
Db 68 TGCTCCAGTAATGGCGGTAGAGCCCGGAGAGGAAATCGGCGCTTACTGGGGGCGCAGGC 9
Qy 168 HisPro 169
Db 8 CATCCT 3

RESULT 3
US-09-796-692A-7011/c
; Sequence 7011, Application US/09796692A
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013512US
; CURRENT APPLICATION NUMBER: US/09/796,692A
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: US 60/186,126
```

```

; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7011
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-796-692A-7011

Alignment Scores:
Pred. No.: 4,34e-55 Length: 310
Score: 541.00 Matches: 97
Percent Similarity: 97.06% Conservatives: 2
Best Local Similarity: 95.10% Mismatches: 3
Query Match: 34.99% Indels: 0
DB: 5 Gaps: 0

US-10-054-313-1 (1-286) x US-09-796-692A-7011 (1-310)

Qy 68 TrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis 87
Db 308 TGGGCCCTTTGTCAGAAATCTGCAAGCCCGGAAGTTTCGAAGGGGCATGAAATCAACAT 249

Qy 88 GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGly 107
Db 248 GGACAAGATCGGAGGCGGAAGCCAGCAGCAGCTCCGTGAGCCACTGGATGGAGATGA 189

Qy 108 HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127
Db 188 CATGAAAGCGCAGAGCCGCTATGCAAGACACATGAAGTCGAGCGTGGAGCGCGGCTCCA 129

Qy 128 ValSerArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCys 147
Db 128 GTTAGCAGAGACAGCTTTCTCATCTGGGAGACTTCCTGCTGTACACTGATGGCTGC 69

Qy 148 CysSerSerAsnGlyArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGly 167
Db 68 TGCCTCCAGTAATGGCGTAGAAGCCCGGAGCAGGAATCGCGCTTTACTGGGCGCGGGC 9

Qy 168 HisPro 169
Db 8 CATCCT 3

RESULT 4
US-60-507-481-9585
; Sequence 9585, Application US/60507481
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION IN ANIMAL
; FILE OF INVENTION: MODLES OF INFLAMMATORY DISEASES
; FILE REFERENCE: AM101084
; CURRENT APPLICATION NUMBER: US/60/507,481
; CURRENT FILING DATE: 2003-10-02

```

```

; NUMBER OF SEQ ID NOS: 210107
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9585
; LENGTH: 629
; TYPE: DNA
; ORGANISM: Canis familiaris
US-60-507-481-9585

Alignment Scores:
Pred. No.: 5,92e-40 Length: 629
Score: 419.00 Matches: 89
Percent Similarity: 84.07% Conservatives: 6
Best Local Similarity: 78.76% Mismatches: 14
Query Match: 27.10% Indels: 4
DB: 7 Gaps: 0

US-10-054-313-1 (1-286) x US-60-507-481-9585 (1-629)

Qy 178 GlyArgGlnThrAsnGln-ArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAl 197
Db 12 GGCTCCAGACCGCAGCAGCAGCAGAAATTCATGCAGCCTGCAAGCCATCGAGCAGCA 71

Qy 197 alysthrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGl 217
Db 72 AAGGCGCCATACACTTTAGAAAGTTAGTTCCTGTATACGACAGCTATGTACTATAATGG 131

Qy 217 yileThrAsnTrpValGlnGlyTyrLysLysAsn-GlyTyrLysThrSerAlaGlyLysG 237
Db 132 TATCATAATTGGGTTCAAGGTTGGAAGAAGAAATGGGTGGAAGAAACAAGCAGGAAAG 191

Qy 237 luValIleAsnLysGluAspPheVal-AlaLeuGluArgLeuThrGlnGlyMetAspIle 256
Db 192 AGGTATCAACAAGGAGGACITTTGGTCCCTGGAGCAGCTCACCCAGGGCATGGACATC 251

Qy 257 GlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeu 276
Db 252 CAGTGGATGCATGTTCTCTGTCATTCGGGATTTACAGGCAATGAAGAAGCTGACAGATTA 311

Qy 277 Ala-ArgGluGlyAlaLysGlnSerGluasp 286
Db 312 GCACAGAGAAGGAGCTAAGCAATCTGAAGAC 342

RESULT 5
US-10-649-136-40/c
; Sequence 40, Application US/10649136
; GENERAL INFORMATION:
; APPLICANT: LAGACE, Robert, E.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BERG, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: ELITRA.025C1
; CURRENT APPLICATION NUMBER: US/10/649,136
; CURRENT FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: 09/596,002
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 40
; LENGTH: 119211
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-10-649-136-40

Alignment Scores:
Pred. No.: 1,59e-18 Length: 119211
Score: 272.00 Matches: 62
Percent Similarity: 52.23% Conservatives: 20
Best Local Similarity: 39.49% Mismatches: 51
Query Match: 17.59% Indels: 24
DB: 6 Gaps: 5

US-10-054-313-1 (1-286) x US-10-649-136-40 (1-119211)

```

QY 136 MetGlyAspPheValValValThrAspGlyCysCysSerSerAsnGlyArgArgLys 155
|||
Db 45182 ATGAGCCAACTTAATCGCTTATACCGACGGCGCTGTAAAGCAATGGAAACAAAGGC 45123
|||
QY 156 ProArgala---GlyIleGlyValTyrr----- 163
Db 45122 GTATCTGCGAGCGGTGGGGCGGTGATTGTCATTATTTCAATGGCGATGAGCGGCATCTG 45063
|||
QY 164 TrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrAsnGln 183
|||
Db 45062 TGGGGC---GGTGAACCT-----GATACGACCAATAAT 45033
|||
QY 184 ArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsn 203
|||
Db 45032 CGCATGGATTGCTGCCATCACAGCTTTGGAGGCAACGCTGCACAG-----ATT 44979
|||
QY 204 LysLeuValLeuTyrrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGln 223
|||
Db 44978 CCTCTGCAACTTTGGACAGATTTCAGGCTATGTAAAGATGGCATTAACCTCAGTGGATTGGC 44919
|||
QY 224 GlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAsp 243
|||
Db 44918 GGTTCGAAGTTCGGTGGTGGAAAAAGCAGATGGCAAGCTGTCTTAATCAAGACCTA 44859
|||
QY 244 PheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMethHisValProGly 263
|||
Db 44858 TGCAACGATTGGATCGCTGACCCCAATCGCATCTTGGTGGCAATGGATCAAGGC 44799
|||
QY 264 HisSerGlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGluGly 280
|||
Db 44798 CATGCTGGTCATGCTGCGCAATGAGATGGCAGACCACTGGCCAAATAAGGC 44748
|||

RESULT 6

US-10-672-787-40/c
; Sequence 40, Application US/10672787
; GENERAL INFORMATION:
; APPLICANT: LAGACE, Robert, E.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BERG, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: ELITRA.025C1
; CURRENT APPLICATION NUMBER: US/10/672,787
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: 09/596,002
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 40
; LENGTH: 119211
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-10-672-787-40

Alignment Scores:
Pred. No.: 1,59e-18 Length: 119211
Score: 272.00 Matches: 62
Percent Similarity: 52.23% Conservative: 20
Best Local Similarity: 39.49% Mismatches: 51
Query Match: 17.59% Indels: 24
DB: 6 Gaps: 5

US-10-054-313-1 (1-286) x US-10-672-787-40 (1-119211)

QY 136 MetGlyAspPheValValValThrAspGlyCysCysSerSerAsnGlyArgArgLys 155
|||
Db 45182 ATGAGCCAACTTAATCGCTTATACCGACGGCGCTGTAAAGCAATGGAAACAAAGGC 45123
|||
QY 156 ProArgala---GlyIleGlyValTyrr----- 163
Db 45122 GTATCTGCGAGCGGTGGGGCGGTGATTGTCATTATTTCAATGGCGATGAGCGGCATCTG 45063
|||
QY 164 TrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrAsnGln 183
|||

Db 45062 TGGGGC---GGTGAACCT-----GATACGACCAATAAT 45033
|||
QY 184 ArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsn 203
|||
Db 45032 CGCATGGATTGCTGCCATCACAGCTTTGGAGGCAACGCTGCACAG-----ATT 44979
|||
QY 204 LysLeuValLeuTyrrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGln 223
|||
Db 44978 CCTCTGCAACTTTGGACAGATTTCAGGCTATGTAAAGATGGCATTAACCTCAGTGGATTGGC 44919
|||
QY 224 GlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAsp 243
|||
Db 44918 GGTTCGAAGTTCGGTGGTGGAAAAAGCAGATGGCAAGCTGTCTTAATCAAGACCTA 44859
|||
QY 244 PheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMethHisValProGly 263
|||
Db 44858 TGCAACGATTGGATCGCTGACCCCAATCGCATCTTGGTGGCAATGGATCAAGGC 44799
|||
QY 264 HisSerGlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGluGly 280
|||
Db 44798 CATGCTGGTCATGCTGCGCAATGAGATGGCAGACCACTGGCCAAATAAGGC 44748
|||

RESULT 7

US-09-806-866A-1
; Sequence 1, Application US/09806866A
; GENERAL INFORMATION:
; APPLICANT: Frazer, Claire M.
; APPLICANT: Hickey, Erin
; APPLICANT: Peterson, Jeremy
; APPLICANT: Tettelin, Hervé
; APPLICANT: Venter, J. Craig
; APPLICANT: Masignani, Vega
; APPLICANT: Galeotti, Cesira
; APPLICANT: Mora, Manroza
; APPLICANT: Ratti, Giulio
; APPLICANT: Scarselli, Maria
; APPLICANT: Scariato, Vincenzo
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisseria Genomic Sequences And Methods Of Their Use
; FILE REFERENCE: CHIR0313
; CURRENT APPLICATION NUMBER: US/09/806,866A
; CURRENT FILING DATE: 2001-09-06
; PRIOR APPLICATION NUMBER: USSN 60/103,794
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: USSN 60/132,068
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/US99/25373
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 1068
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 36471
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
US-09-806-866A-1

Alignment Scores:
Pred. No.: 2,72e-14 Length: 36471
Score: 229.50 Matches: 54
Percent Similarity: 48.73% Conservative: 23
Best Local Similarity: 34.18% Mismatches: 56
Query Match: 14.84% Indels: 25
DB: 5 Gaps: 5

US-10-054-313-1 (1-286) x US-09-806-866A-1 (1-36471)

QY 136 MetGlyAspPheValValValThrAspGlyCysCysSerSerAsnGlyArgArgLys 155
|||
Db 30559 ATGAACCAACCCGTTTACCTTTTACCCGCGCGCGTCAAGGCAAT----- 30606
|||
QY 156 ProArgAlaGlyIleGlyValTyrrTrpGlyProGlyHisProLeuAsnValGlyIleArg 175
|||

Db 30607 CCCGCGGGGC-----GGCTGGGC-----GTGTTAATGCG 30639
Qy 176 LeuProGlyArgGln-----ThraSnglnArgAla 185
Db 30640 TACGGTACCCAGCAAGAAAGACTTTTCGGCGCGGAGCGCAACACCAACCGCATG 30699
Qy 186 GluileHisAlaAalaCysLysAlaileGluGlnAlaLysThrGlnAsnIleAsnLysLeu 205
Db 30700 GAACGTACTCGCTCATGAGGACTGAAATCGCTCAAAACCGCG-----TGACCGTC 30753
Qy 206 ValLeuTyThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrp 225
Db 30754 ATCATCTGCACCGACTCGCAATACGTCAAAATGCGTGAAGAACTGGATACACGGTTGG 30813
Qy 226 LysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheVal 245
Db 30814 AGCCCAACGGCTGGAACACCGCTCCAAACAGCCCGCAAAACGAGCATTTGTGAAA 30873
Qy 246 AlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSer 265
Db 30874 GAACTCGACGCTCTAGTCGCGCGCATCAAGTCAGTTGGCTGGAAGGACACGCG 30933
Qy 266 GlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGluGlyAlaLysGln 283
Db 30934 GGACACGCGCAACGAAACGCGCGCAGCATTTGGCAAAACCGTGGCGCGACGCG 30987

RESULT 8

US-09-806-866A-1068
; Sequence 1068, Application US/09806866A
; GENERAL INFORMATION:
; APPLICANT: Frazer, Claire M.
; APPLICANT: Hickey, Erin
; APPLICANT: Peterson, Jeremy
; APPLICANT: Tettelin, Hervé
; APPLICANT: Venter, J. Craig
; APPLICANT: Massignani, Vega
; APPLICANT: Galeotti, Cesira
; APPLICANT: Mora, Manroza
; APPLICANT: Ratti, Giulio
; APPLICANT: Scarcelli, Maria
; APPLICANT: Scarlato, Vincenzo
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisseria Genomic Sequences And Methods Of Their Use
; FILE REFERENCE: CHIR0313
; CURRENT APPLICATION NUMBER: US/09/806,866A
; PRIOR FILING DATE: 2001-09-06
; PRIOR APPLICATION NUMBER: USSN 60/103,794
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: USSN 60/132,068
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/US99/25373
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 1068
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1068
; LENGTH: 2242716
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
US-09-806-866A-1068

Alignment Scores:
Pred. No.: 2,68e-11 Length: 2242716
Score: 229.50 Matches: 54
Percent Similarity: 48.73% Conservative: 23
Best Local Similarity: 34.18% Mismatches: 56
Query Match: 14.84% Indels: 25
DB: 5 Gaps: 5

US-10-054-313-1 (1-286) x US-09-806-866A-1068 (1-2242716)

Qy 136 MetGlyAspPheValValValValTyThrAspGlyCysCysSerSerAsnGlyArgArgLys 155
Db 1651985 ATGAACCAACCGTTTACCTTTACACCGACGGCGGTGCAAAAGGCAAT-----1652032
Qy 156 ProArgAlaGlyIleGlyValTyTrpGlyProGlyHisProLeuAsnValGlyIleArg 175
Db 1652033 CCGCGCGCGGC-----GGCTGGGC-----GTGTTAATGCGC 1652065
Qy 176 LeuProGlyArgGln-----ThraSnglnArgAla 185
Db 1652066 TACGGTAGCCACCAAGAAAGACTTTTCGGCGCGGAGCGCAACACCAACCGCATG 1652125
Qy 186 GluileHisAlaAalaCysLysAlaileGluGlnAlaLysThrGlnAsnIleAsnLysLeu 205
Db 1652126 GAACGTACTCGCTCATGAGGACTGAAATCGCTCAAAACCGCG-----TGCACCGTC 1652179
Qy 206 ValLeuTyThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrp 225
Db 1652180 ATCATCTGCACCGACTCGCAATACGTCAAAATGCGTGAAGAACTGGATACACGGTTGG 1652239
Qy 226 LysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheVal 245
Db 1652240 AACGCAACCGCTGGAACACCGCTCCAAACAGCCCGTCAAAACGAGCATTTGTGAAA 1652299
Qy 246 AlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSer 265
Db 1652300 GAACTCGACGCTCTAGTCGCGCGCATCAAGTCAGTTGGCTGGAAGGACACGCG 1652359
Qy 266 GlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGluGlyAlaLysGln 283
Db 1652360 GGACACGCGCAACGAAACGCGCGCAGCATTTGGCAAAACCGTGGCGCGACGCG 1652413

RESULT 9

US-09-897-516A-658
; Sequence 658, Application US/09897516A
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215,161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8415
; SEQ ID NO 658
; LENGTH: 1733
; TYPE: DNA
; ORGANISM: Xenorhabdus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (874)..(1335)
; OTHER INFORMATION:
US-09-897-516A-658

Alignment Scores:
Pred. No.: 1.8e-14 Length: 1733
Score: 212.50 Matches: 56
Percent Similarity: 47.10% Conservative: 17
Best Local Similarity: 36.13% Mismatches: 71
Query Match: 13.75% Indels: 11
DB: 5 Gaps: 4

US-10-054-313-1 (1-286) x US-09-897-516A-658 (1-1733)

Qy 130 ArgAspThrPheSerTyMetGlyAspPheValValTyThrAspGlyCysCysSer 149
Db 1652413 ATGAACCAACCGTTTACCTTTACACCGACGGCGGTGCAAAAGGCAAT-----1652032

Db 856 AGGAGAGCTACCAAAATATGAATAACACAGGTAGAAATTTTACCGGATGGTTCCTGCCTC 915
Qy 150 SerAsnGlyArgGlyProArgAlaGlyIleGlyVal-----TyrTrpGlyPro 166
Db 916 GGTAAT-----CCGGTCCCGCGGATATGGTGTGTACTTCGTATATAACAAAAG 966
Qy 167 GlyHisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGlu 186
Db 967 GAAAAAACCCTTAATAAAGTTAT-----TTCGGTACACCAATAAACCGGATGGAG 1017
Qy 187 IleHisAlaAlaCysGlyAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuVal 206
Db 1018 CTAATGGCGCTATCCCGCGTGAAGCCCTTAAAGAGCCA-----TGCTCAATCATT 1071
Qy 207 LeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLys 226
Db 1072 CTTACCACTGACAGCAATATGTTCTGCAGGAAATACCAGTGGATACATACTGGAAA 1131
Qy 227 LysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAla 246
Db 1132 AAACGGGGCTGGAAAAAGCTGATAAATCCCGGTGATCAATGTTGATCTCTGGCAACGG 1191
Qy 247 LeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGly 266
Db 1192 CTGGATPAAAGCATTCTCTCAGGAATCGAATGGCATTGGTCAAGGGGATACCGGA 1251
Qy 267 PheIleGlyAsnGluAlaAspArgLeuAlaArgGluGlyAla 281
Db 1252 CATCGTGAACGAATGTGTGATGAATCGCGCCGTCGACGGCT 1296
RESULT 10
US-09-897-516A-659/c
; Sequence 659, Application US/09897516A
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516A
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215,161
; NUMBER OF SEQ ID NOS: 8415
; SEQ ID NO 659
; LENGTH: 1733
; TYPE: DNA
; ORGANISM: Xenorhabdus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (925)..(1674)
; OTHER INFORMATION:
US-09-897-516A-659
Alignment Scores:
Pred. No.: 1.8e-14 Length: 1733
Score: 212.50 Matches: 56
Percent Similarity: 47.10% Conservative: 17
Best Local Similarity: 36.13% Mismatches: 71
Query Match: 13.75% Indels: 11
DB: 5 Gaps: 4

US-10-054-313-1 (1-286) x US-09-897-516A-659 (1-1733)

Qy 130 ArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCysSer 149
Db 878 AGGAAGAGTCTACCAATATGATAAATACAGGTAGAAATTTTACCGGATGGTTCCTGCCTC 819

Qy 150 SerAsnGlyArgGlyProArgAlaGlyIleGlyVal-----TyrTrpGlyPro 166
Db 818 GGTAAT-----CCGGTCCCGCGGATATGGTGTGTACTTCGTATATAACAAAAG 768
Qy 167 GlyHisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGlu 186
Db 767 GAAAAAACCCTTAATAAAGTTAT-----TTCGGTACACCAATAAACCGGATGGAG 717
Qy 187 IleHisAlaAlaCysGlyAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuVal 206
Db 716 CTAATGGCGCTATCCCGCGTGAAGCCCTTAAAGAGCCA-----TGCTCAATCATT 663
Qy 207 LeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLys 226
Db 662 CTTACCACTGACAGCAATATGTTCTGCAGGAAATACCAGTGGATACATACTGGAAA 503
Qy 227 LysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAla 246
Db 502 AAACGGGGCTGGAAAAAGCTGATAAATCCCGGTGATCAATGTTGATCTCTGGCAACGG 543
Qy 247 LeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGly 266
Db 542 CTGGATPAAAGCATTCTCTCAGGAATCGAATGGCATTGGTCAAGGGGATACCGGA 483
Qy 267 PheIleGlyAsnGluAlaAspArgLeuAlaArgGluGlyAla 281
Db 482 CATCGTGAACGAATGTGTGATGAATCGCGCCGTCGACGGCT 438
RESULT 11
US-09-897-516A-660/c
; Sequence 660, Application US/09897516A
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516A
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215,161
; NUMBER OF SEQ ID NOS: 8415
; SEQ ID NO 660
; LENGTH: 1733
; TYPE: DNA
; ORGANISM: Xenorhabdus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(352)
; OTHER INFORMATION:
US-09-897-516A-660
Alignment Scores:
Pred. No.: 1.8e-14 Length: 1733
Score: 212.50 Matches: 56
Percent Similarity: 47.10% Conservative: 17
Best Local Similarity: 36.13% Mismatches: 71
Query Match: 13.75% Indels: 11
DB: 5 Gaps: 4

US-10-054-313-1 (1-286) x US-09-897-516A-660 (1-1733)

Qy 130 ArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCysSer 149
Db 878 AGGAAGAGTCTACCAATATGATAAATACAGGTAGAAATTTTACCGGATGGTTCCTGCCTC 819
Qy 150 SerAsnGlyArgGlyProArgAlaGlyIleGlyVal-----TyrTrpGlyPro 166

```
Db 818 GGTAAAT-----CCGGTCCCGCGGATATGTTGTTTACTTCTGTTATAAACAAG 768
Qy 167 GlyHisProLeuAsnValGlyLeuArgLeuProGlyArgGlnThrAsnGlnArgAlaGlu 186
Db 767 GAAAAAACCCTTAAATAAGTTAT-----TTCCGTACAAACCAATAACCGGATGAG 717
Qy 187 IleHisAlaAlaCysLeuAlaIleGlnGlnAlaLysThrGlnAsnIleAsnLysLeuVal 206
Db 716 CTAATGGCGGCTATCGCGCGCTGAAGCCCTTAAAGCCA-----TGCTCAATCAT 663
Qy 207 LeuTyThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTyrLys 226
Db 662 CTTACCACTGACAGCAATATGCTGTCAGGAATATCCAGTGATACATAAATGGAAA 603
Qy 227 LysAsnGlyTyrLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValala 246
Db 602 AAACGGGGCTGAAAAAGCTGATAAATCCCGGTGATCAATGTTGATCTGTGCAACGG 543
Qy 247 LeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGly 266
Db 542 CTGGATAAAGCCATTCTCTCCACGAATCGATGGCATTGGGTCAAGGGGCATACCGGA 483
Qy 267 PheIleGlyAsnGluAlaAspArgLeuAlaArgGluGlyAla 281
Db 482 CATCGTAAACGAATTGTGTGATGAATCGCCCGCTGCAGCGGCT 438
RESULT 12
US-10-684-141-41
; Sequence 41, Application US/10684141
; GENERAL INFORMATION:
; APPLICANT: Furusawa, Mitsuru
; TITLE OF INVENTION: METHOD AND SYSTEM FOR RAPIDLY CONFERRING
; FILE REFERENCE: 690116.401
; CURRENT APPLICATION NUMBER: US/10/684,141
; PRIOR FILING DATE: 2003-10-10
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 1592
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-684-141-41
Alignment Scores:
Pred. No.: 4,71e-14 Length: 1592
Score: 208.50 Matches: 52
Percent Similarity: 47.89% Conservative: 16
Best Local Similarity: 36.62% Mismatches: 69
Query Match: 13.49% Indels: 5
DB: 6 Gaps: 2
US-10-054-313-1 (1-286) x US-10-684-141-41 (1-1592)
Qy 140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly 159
Db 1088 GTGAAATTTTCCAGATGTTGTTGTTCTGGCAATCCAGGACCTGGGGGTTAGCGGCT 1147
Qy 160 IleGlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArg 179
Db 1148 ATTTTACGATATCGCGGACCGGAGAAAAACCTTTAGCGCTGGCTACACC-----CGC 1198
Qy 180 GlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThr 199
Db 1199 ACCACCAACACCGTATGAGTGTATGCGCGCTATGTCGCGCTGAGCGGCTTAAAGAA 1258
Qy 200 GlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThr 219
Db 1259 CAT-----TGCGAAGTCATTTGATACCGACACCGCATATGTCGCGCGGATACCC 1312
Qy 220 AsnTrpValGlnGlyTyrLysLysAsnGlyTyrLysThrSerAlaGlyLysGluValle 239
Db 1313 CAGTGGATCCATPACTGGAAAAAACCGTGTGAAAAACCGCAGACAAAAAACCAAGTAAA 1372
```

```
Qy 240 AsnLysGluAspPheValalaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMet 259
Db 1373 AATGTCGATCTCTGGCAACGTTCTTGATCTGGGCGCAGCATCAATCAATCGGAA 1432
Qy 260 HisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGlu 279
Db 1433 TGGGTTAAAGGCCATCCGCGACACCCGGAACGAAACGCTGTGATGAATCGCTCGTCC 1492
Qy 280 GlyAla 281
Db 1493 GCGGCG 1498
RESULT 13
US-10-297-465B-1/c
; Sequence 1, Application US/10297465B
; GENERAL INFORMATION:
; APPLICANT: Simpson, Andrew
; APPLICANT: Reinach, Fernando
; APPLICANT: Setubal, Joao
; APPLICANT: Medanis, Joao
; APPLICANT: Arruda, Paulo
; TITLE OF INVENTION: Isolated Genome of Xylella Fastidiosa and Uses Thereof
; FILE REFERENCE: FAPESP 202 US (10213376)
; CURRENT APPLICATION NUMBER: US/10/297,465B
; CURRENT FILING DATE: 2003-08-25
; PRIOR APPLICATION NUMBER: PCT/IB01/01618
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,906
; PRIOR FILING DATE: 2001-06-17
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465B-1
Alignment Scores:
Pred. No.: 5,84e-07 Length: 2731748
Score: 194.50 Matches: 51
Percent Similarity: 46.90% Conservative: 17
Best Local Similarity: 35.17% Mismatches: 54
Query Match: 12.58% Indels: 23
DB: 6 Gaps: 4
US-10-054-313-1 (1-286) x US-10-297-465B-1 (1-2731748)
Qy 143 TyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIleGlyVal 162
Db 2049147 TACACGACGGTCTTGTCTTGTGTAATCCT----- 2049118
Qy 163 TyrTrpGlyProGly-----HisProLeuAsnValGlyIleArgLeu 176
Db 2049117 -----GGTCTGTTGGCTGTTGTTGCGTTATATAAACAATGAGAAGGAGTTA 2049064
Qy 177 ProGlyArgGln-----ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAla 193
Db 2049063 GTTGGTGGTGAGCTTGATACGACCAATAATCGAATCGAGCTGATGCTGCGATTATGGCT 2049004
Qy 194 IleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValIleTyrThrAspSerMetPhe 213
Db 2049003 CTGGAG-----CGGTGAGTGAGCCTTGTGATCAATCAAGCTTCATCTGTCGAGTAT 2048950
Qy 214 ThrIleAsnGlyIleThrAsnTrpValGlnGlyTyrLysLysAsnGlyTyrLysThrSer 233
Db 2048949 GTCCGCAAGGATCACTGATGATGTCGCGTGGGTACGCGCTGGTGGTGGTGGTGGTGG 2048890
Qy 234 AlaGlyLysGluValIleAsnLysGluAspPheValalaLeuGluArgLeuThrGlnGly 253
Db 2048889 GCGCGTACCCCGGTCAAGATCGTATGATGATGAGCGGCTTGTGTCGCGCTACGACGCG 2048830
Qy 254 MetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluAla 273
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 17, 2003, 19:17:25 ; Search time 3541 Seconds
(without alignments)
2678.577 Million cell updates/sec

Title: US-10-054-313-1

Perfect score: 1546

Sequence: 1 MSWLLFLAHRVALAALPCRR.....FIGNEADRLAREGAKQSED 286

Scoring table: BLOSUM62

| |
|---------------------------|
| Xgapop 10.0 , Xgapext 0.5 |
| Ygapop 10.0 , Ygapext 0.5 |
| Fgapop 6.0 , Fgapext 7.0 |
| Delop 6.0 , Delext 7.0 |

Searched: 33363688 seqs, 16581889874 residues

Total number of hits satisfying chosen parameters: 66727376

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US10054313/runat 17122003 150745 24277/app query.fasta 1.455
-DB=Pending Patents NA Main -OFFWT=fastap -SUFFIX=rnrm -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -SPART=1 -END=1 -NATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054313 @cgn2_1 3914 -runat 17122003 150745 24277 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Pending Patents NA Main:

| | |
|---|--|
| 1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq.* | 29: /cgn2_6/ptodata/2/pna/US096D_COMB.seq.* |
| 2: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq.* | 30: /cgn2_6/ptodata/2/pna/US096E_COMB.seq.* |
| 3: /cgn2_6/ptodata/2/pna/US06_COMB.seq.* | 31: /cgn2_6/ptodata/2/pna/US097A_COMB.seq.* |
| 4: /cgn2_6/ptodata/2/pna/US07_COMB.seq.* | 32: /cgn2_6/ptodata/2/pna/US097B_COMB.seq.* |
| 5: /cgn2_6/ptodata/2/pna/US080_COMB.seq.* | 33: /cgn2_6/ptodata/2/pna/US097C_COMB.seq.* |
| 6: /cgn2_6/ptodata/2/pna/US081_COMB.seq.* | 34: /cgn2_6/ptodata/2/pna/US098A_COMB.seq.* |
| 7: /cgn2_6/ptodata/2/pna/US082_COMB.seq.* | 35: /cgn2_6/ptodata/2/pna/US098B_COMB.seq.* |
| 8: /cgn2_6/ptodata/2/pna/US083_COMB.seq.* | 36: /cgn2_6/ptodata/2/pna/US098C_COMB.seq.* |
| 9: /cgn2_6/ptodata/2/pna/US084_COMB.seq.* | 37: /cgn2_6/ptodata/2/pna/US098D_COMB.seq.* |
| 10: /cgn2_6/ptodata/2/pna/US085_COMB.seq.* | 38: /cgn2_6/ptodata/2/pna/US099A_COMB.seq.* |
| 11: /cgn2_6/ptodata/2/pna/US086_COMB.seq.* | 39: /cgn2_6/ptodata/2/pna/US099B_COMB.seq.* |
| 12: /cgn2_6/ptodata/2/pna/US087_COMB.seq.* | 40: /cgn2_6/ptodata/2/pna/US099C_COMB.seq.* |
| 13: /cgn2_6/ptodata/2/pna/US088_COMB.seq.* | 41: /cgn2_6/ptodata/2/pna/US099D_COMB.seq.* |
| 14: /cgn2_6/ptodata/2/pna/US089_COMB.seq.* | 42: /cgn2_6/ptodata/2/pna/US099E_COMB.seq.* |
| 15: /cgn2_6/ptodata/2/pna/US090_COMB.seq.* | 43: /cgn2_6/ptodata/2/pna/US099F_COMB.seq.* |
| 16: /cgn2_6/ptodata/2/pna/US091_COMB.seq.* | 44: /cgn2_6/ptodata/2/pna/US100A_COMB.seq.* |
| 17: /cgn2_6/ptodata/2/pna/US092A_COMB.seq.* | 45: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.* |
| 18: /cgn2_6/ptodata/2/pna/US092B_COMB.seq.* | 46: /cgn2_6/ptodata/2/pna/US101A_COMB.seq.* |
| 19: /cgn2_6/ptodata/2/pna/US093A_COMB.seq.* | 47: /cgn2_6/ptodata/2/pna/US101B_COMB.seq.* |
| 20: /cgn2_6/ptodata/2/pna/US093B_COMB.seq.* | 48: /cgn2_6/ptodata/2/pna/US102A_COMB.seq.* |
| 21: /cgn2_6/ptodata/2/pna/US094_COMB.seq.* | 49: /cgn2_6/ptodata/2/pna/US102B_COMB.seq.* |
| 22: /cgn2_6/ptodata/2/pna/US095A_COMB.seq.* | 50: /cgn2_6/ptodata/2/pna/US103A_COMB.seq.* |
| 23: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.* | 51: /cgn2_6/ptodata/2/pna/US103B_COMB.seq.* |
| 24: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.* | 52: /cgn2_6/ptodata/2/pna/US104A_COMB.seq.* |
| 25: /cgn2_6/ptodata/2/pna/US095D_COMB.seq.* | 53: /cgn2_6/ptodata/2/pna/US104B_COMB.seq.* |
| 26: /cgn2_6/ptodata/2/pna/US096A_COMB.seq.* | 54: /cgn2_6/ptodata/2/pna/US6000_COMB.seq.* |
| 27: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.* | 55: /cgn2_6/ptodata/2/pna/US6001_COMB.seq.* |
| 28: /cgn2_6/ptodata/2/pna/US096C_COMB.seq.* | 56: /cgn2_6/ptodata/2/pna/US6002_COMB.seq.* |
| | 57: /cgn2_6/ptodata/2/pna/US6003_COMB.seq.* |
| | 58: /cgn2_6/ptodata/2/pna/US6004_COMB.seq.* |
| | 59: /cgn2_6/ptodata/2/pna/US6005_COMB.seq.* |
| | 60: /cgn2_6/ptodata/2/pna/US6006_COMB.seq.* |
| | 61: /cgn2_6/ptodata/2/pna/US6007_COMB.seq.* |
| | 62: /cgn2_6/ptodata/2/pna/US6008_COMB.seq.* |
| | 63: /cgn2_6/ptodata/2/pna/US6009_COMB.seq.* |
| | 64: /cgn2_6/ptodata/2/pna/US6010_COMB.seq.* |
| | 65: /cgn2_6/ptodata/2/pna/US6011_COMB.seq.* |
| | 66: /cgn2_6/ptodata/2/pna/US6012_COMB.seq.* |
| | 67: /cgn2_6/ptodata/2/pna/US6013_COMB.seq.* |
| | 68: /cgn2_6/ptodata/2/pna/US6014_COMB.seq.* |
| | 69: /cgn2_6/ptodata/2/pna/US6015_COMB.seq.* |
| | 70: /cgn2_6/ptodata/2/pna/US6016_COMB.seq.* |
| | 71: /cgn2_6/ptodata/2/pna/US6017_COMB.seq.* |
| | 72: /cgn2_6/ptodata/2/pna/US6018_COMB.seq.* |
| | 73: /cgn2_6/ptodata/2/pna/US6019_COMB.seq.* |
| | 74: /cgn2_6/ptodata/2/pna/US6020_COMB.seq.* |
| | 75: /cgn2_6/ptodata/2/pna/US6021_COMB.seq.* |
| | 76: /cgn2_6/ptodata/2/pna/US6022_COMB.seq.* |
| | 77: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq.* |
| | 78: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq.* |
| | 79: /cgn2_6/ptodata/2/pna/US6024_COMB.seq.* |
| | 80: /cgn2_6/ptodata/2/pna/US6025_COMB.seq.* |
| | 81: /cgn2_6/ptodata/2/pna/US6026_COMB.seq.* |
| | 82: /cgn2_6/ptodata/2/pna/US6027_COMB.seq.* |
| | 83: /cgn2_6/ptodata/2/pna/US6028_COMB.seq.* |
| | 84: /cgn2_6/ptodata/2/pna/US6029_COMB.seq.* |
| | 85: /cgn2_6/ptodata/2/pna/US6030_COMB.seq.* |
| | 86: /cgn2_6/ptodata/2/pna/US6031_COMB.seq.* |
| | 87: /cgn2_6/ptodata/2/pna/US6032_COMB.seq.* |
| | 88: /cgn2_6/ptodata/2/pna/US6033_COMB.seq.* |
| | 89: /cgn2_6/ptodata/2/pna/US6034_COMB.seq.* |
| | 90: /cgn2_6/ptodata/2/pna/US6035_COMB.seq.* |
| | 91: /cgn2_6/ptodata/2/pna/US6036_COMB.seq.* |
| | 92: /cgn2_6/ptodata/2/pna/US6037_COMB.seq.* |
| | 93: /cgn2_6/ptodata/2/pna/US6038_COMB.seq.* |
| | 94: /cgn2_6/ptodata/2/pna/US6039_COMB.seq.* |
| | 95: /cgn2_6/ptodata/2/pna/US6040_COMB.seq.* |
| | 96: /cgn2_6/ptodata/2/pna/US6041_COMB.seq.* |
| | 97: /cgn2_6/ptodata/2/pna/US6042_COMB.seq.* |
| | 98: /cgn2_6/ptodata/2/pna/US6043_COMB.seq.* |
| | 99: /cgn2_6/ptodata/2/pna/US6044_COMB.seq.* |
| | 100: /cgn2_6/ptodata/2/pna/US6045_COMB.seq.* |
| | 101: /cgn2_6/ptodata/2/pna/US6046_COMB.seq.* |

102: /cgn2_6/prodata/2/pna/us6047_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 1526 | 98.7 | 1526 | 30 | US-09-698-010-12700 |
| 2 | 1526 | 98.7 | 1526 | 30 | US-09-698-013-6258 |
| 3 | 1526 | 98.7 | 1526 | 30 | US-09-698-097-12450 |
| 4 | 1526 | 98.7 | 1526 | 31 | US-09-700-000-5074 |
| 5 | 1526 | 98.7 | 1526 | 31 | US-09-198-846-12889 |
| 6 | 1526 | 98.7 | 1526 | 33 | US-09-770-171-1164 |
| 7 | 1526 | 98.7 | 1526 | 33 | US-09-785-276A-30281 |
| 8 | 1526 | 98.7 | 1526 | 33 | US-09-357-930-30281 |
| 9 | 1523 | 98.5 | 1523 | 31 | US-09-763-233-41 |
| 10 | 1523 | 98.5 | 1523 | 31 | US-09-115-639-32 |
| 11 | 1523 | 98.5 | 1523 | 65 | US-60-115-639-32 |
| 12 | 1522 | 98.4 | 1522 | 9 | US-10-170-235-15341 |
| 13 | 1522 | 98.4 | 1522 | 9 | US-10-262-511-85 |
| 14 | 1522 | 98.4 | 1522 | 9 | US-10-262-511A-85 |
| 15 | 1522 | 98.4 | 1522 | 1 | US-10-262-511A-85 |
| 16 | 1522 | 98.4 | 1522 | 1 | US-10-262-511A-85 |
| 17 | 1522 | 98.4 | 1522 | 47 | US-10-172-118-841 |
| 18 | 1522 | 98.4 | 1522 | 47 | US-10-342-887-841 |
| 19 | 1522 | 98.4 | 1522 | 71 | US-60-172-373-7744 |
| 20 | 1522 | 98.4 | 1522 | 71 | US-60-172-373-7744 |
| 21 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 22 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 23 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 24 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 25 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 26 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 27 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 28 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 29 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 30 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 31 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 32 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 33 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 34 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 35 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 36 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 37 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 38 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 39 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 40 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 41 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 42 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 43 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 44 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |
| 45 | 1522 | 98.4 | 1522 | 82 | US-60-278-258-8522 |

ALIGNMENTS

RESULT 1
US-09-698-010-12700
; Sequence 12700, Application US/09698010
; GENERAL INFORMATION:
; APPLICANT: Williamson, Mark
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: 1600.2029-001
; CURRENT APPLICATION NUMBER: US/09/698,010
; CURRENT FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: 60/162,358
; PRIOR FILING DATE: 1999-10-29

; NUMBER OF SEQ ID NOS: 15684
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12700
; LENGTH: 1652
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-698-010-12700

Alignment Scores:
Pred. No.: 1.33e-119 Length: 1552
Score: 1526.00 Matches: 282
Percent Similarity: 99.30% Conservative: 2
Best Local Similarity: 98.60% Mismatches: 2
Query Match: 98.71% Indels: 0
DB: 30 Gaps: 0

US-10-054-313-1 (1-286) x US-09-698-010-12700 (1-1652)

Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 112 ATGAGCTGGCTTCTGTTCTGCTGCCACAGAGTCCCTTGGCCGCTTGGCCGCGC 171

Qy 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgArgGlyArgGlyThrGlyValPhe 40
Db 172 GGCTCTCGCGGGTTCGGGATGTTCTATGCGTGGAGGGCGGCGGAGCCGGTCTTT 231

Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db 232 CTGACCTGGATGAGTGCAGAGCAGAGTGGACCGTCTTCTGCTGCCAGATTAAAGAAG 231

Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 292 TTTGCCACAGAGATGAGGCTCGGCTTGTTCAGGAAATCTGCAAGCCCGGAGATTCA 351

Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 352 GAAGGCGATGAAATCAACATGACAGAGATCGAGGCGGAGGCGGAGGCGGAGGCGG 411

Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTrpAlaLysHisMetLysPro 120
Db 412 GAGCCACTGGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 471

Qy 121 SerValGluProAlaProProValSerArgAspThrPheSerTrpMetGlyAspPheVal 140
Db 472 AGCTGGAGCGGCGCTCCAGTTAGCAGAGACACAGTTTCTTACATGGAGAGCTTCGTC 531

Qy 141 ValValTrpThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160
Db 532 GTCGCTACACTGATGGCTGCTGCTCCAGTAATGGCGTAGAAGGCGGAGGCGGAGGAG 591

Qy 161 GlyValTrpTrpGlyProGlyHisProLeuAsnValGlyLeuArgLeuProGlyArgGln 180
Db 592 GCGCTTTACTGGGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 651

Qy 181 ThrAsnGlnArgAlaGluLeuHisAlaAlaCysLysAlaLeuGluGlnAlaLysThrGln 200
Db 652 ACAAACCAAGAGCGGAAATTCATGAGCTGCTGAGGCTGCAAGGCGGAGGCGGAGGCG 711

Qy 201 AsnLeuAsnLysLeuValLeuTrpThrAspSerMetPheThrLeuAsnGlyLeuThrAsn 220
Db 712 AACATCAATAAAGCTGGTCTGTATACAGACAGATGTTTACGATAAATGGTATAACTAAC 771

Qy 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValLeuAsn 240
Db 772 TGGGTTCAAGGTTGGAAGAAATAATGGTGGAGACAGATGCGAGGAGGAGGATGATCAAC 831

Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspLysLeuTrpMetHis 260
Db 832 AAGAGGAGCTTTGTGGCACTGGAGAGGCTTACCCAGGCGGATGGACATTCAGTGGATGCAT 891

Qy 261 ValProGlyHisSerGlyPheLeuGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db 892 GTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 951

QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
 Db 352 GAAGGGCATGAAATCAACATGACACAGAAATCGAGCGGAAAGCCAGCAAGCACTCCGT 411
 QY 101 GluProLeuAspGlyHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 120
 Db 412 GAGCCACTGATGGATGAGATGACATGAAAGCGCAGAGCCGTATGCAAGCACATGAGCGC 471
 QY 121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
 Db 472 AGCGTGGAGCGCGCTCCAGTTAGCAGACACAGCTTTTCTTACATGGAGACTTCGTG 531
 QY 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProAlaGlyLys 160
 Db 532 GTCGTCTACACTGATGGCTGCTCTCCAGTAATGGCGGTAGAGCGCGGCGAGCAAGCAATC 591
 QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLysArgLeuArgGln 180
 Db 592 GCGGTTTACTGGGCGCAGCGCATCTTTAAATGTAGGCATTAGACTTCTTGGCGCGCAG 651
 QY 181 ThrAsnGlnArgAlaGluLysHisAlaAlaCysLysAlaLysGlnAlaLysThrGln 200
 Db 652 ACAACCAAGAGCGGAAATTCATGAGCGCTGCAAGCCATTGAACCAAGCAAGCACTCAA 711
 QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyLysThrAsn 220
 Db 712 AACATCAATAACTGGTCTCTGTATACAGACAGTATGTTACGATAAATGGTAACTAAC 771
 QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValLys 240
 Db 772 TGGGTTCAAGTTTGGAGAAATGGGTGGAAGCAAGTGCAGGAAAGAGGTGATCAAC 831
 QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
 Db 832 AAAGAGACTTTGGCCTGGAGAGCTTACCAGGGATGGACATTGATGGAGTGCAT 891
 QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluLysAlaAspArgLeuAlaArgGluGly 280
 Db 892 GTTCTGTCATTCGGGATTTATAGGCAATGAGAGCTGACAGATTAGCCAGAGAGGA 951
 QY 281 AlalysGlnSerGluAsp 286
 Db 952 GCTAAACAATCGGAAGAC 969

RESULT 4

US-09-700-000-5074
 ; Sequence 5074, Application US/09700000
 ; GENERAL INFORMATION:
 ; APPLICANT: Richardson, Jennifer
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
 ; FILE REFERENCE: 1600, 2022-001
 ; CURRENT APPLICATION NUMBER: US/09/700,000
 ; PRIOR FILING DATE: 2000-10-30
 ; PRIOR FILING DATE: 1999-10-25
 ; NUMBER OF SEQ ID NOS: 2171
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5074
 ; LENGTH: 1652
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-700-000-5074

late

Alignment Scores:
 Pred. No.: 1.33e-119 Length: 1652
 Score: 1526.00 Matches: 282
 Percent Similarity: 99.30% Conservative: 2
 Best Local Similarity: 98.60% Mismatches: 2
 Query Match: 98.71% Indels: 0
 DB: 31 Gaps: 0

RESULT 5

US-10-198-846-12889
 ; Sequence 12889, Application US/10198846
 ; GENERAL INFORMATION:
 ; APPLICANT: Lillie, James
 ; APPLICANT: Xu, Yongyao
 ; APPLICANT: Wang, Youzhen
 ; APPLICANT: Steinmann, Kathleen
 ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
 ; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
 ; TITLE OF INVENTION: THERAPY OF BREAST CANCER

US-10-054-313-1 (1-286) x US-09-700-000-5074 (1-1652)
 QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuPheProCysArgArg 20
 Db 112 ATGAGCTGGCTTCTGTCTCTGGCCACAGAGTGGCTTGGCGCTTGGCTTGGCGCGC 171
 QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
 Db 172 GGCCTCTCGCGGTTCGGGATGTTCTATCGCTGAGGAGGGCGCCAAAGACCGGGGTCTTT 231
 QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaAspPheLysLys 60
 Db 232 CTGACCTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 291
 QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
 Db 292 TTTGCCACAGAGGATGAGGCTGGGCTTTGTCTGAGGAAATCTGCAAGCCGCGAAGTTTCA 351
 QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
 Db 352 GAAGGGCATGAAATCAACATGAGCAAGAAATCGAGCGGAAAGCCAGCAAGCACTCCGT 411
 QY 101 GluProLeuAspGlyAspGlyHisGlyGlnGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
 Db 412 GAGCCACTGATGGATGAGATGACATGAAAGCGCAGAGCGGTATGCAAGCACATGAGCGC 471
 QY 121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
 Db 472 AGCGTGGAGCGCGCTCCAGTTAGCAGACAGCTTTTCTTACATGGAGACTTCGTG 531
 QY 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProAlaGlyLys 160
 Db 532 GTCGTCTACACTGATGGCTGCTCTCCAGTAATGGCGGTAGAGCGCGGCGAGCAAGCAATC 591
 QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLysArgLeuProGlyArgGln 180
 Db 592 GCGGTTTACTGGGCGCAGCGCATCTTTAAATGTAGGCATTAGACTTCTTGGCGCGCAG 651
 QY 181 ThrAsnGlnArgAlaGluLysHisAlaAlaCysLysAlaLysGlnAlaLysThrGln 200
 Db 652 ACAACCAAGAGCGGAAATTCATGAGCGCTGCAAGCCATTGAACCAAGCAAGCACTCAA 711
 QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyLysThrAsn 220
 Db 712 AACATCAATAACTGGTCTCTGTATACAGACAGTATGTTACGATAAATGGTAACTAAC 771
 QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValLys 240
 Db 772 TGGGTTCAAGTTTGGAGAAATGGGTGGAAGCAAGTGCAGGAAAGAGGTGATCAAC 831
 QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
 Db 832 AAAGAGACTTTGGCCTGGAGAGCTTACCAGGGATGGACATTGATGGAGTGCAT 891
 QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluLysAlaAspArgLeuAlaArgGluGly 280
 Db 892 GTTCTGTCATTCGGGATTTATAGGCAATGAGAGCTGACAGATTAGCCAGAGAGGA 951
 QY 281 AlalysGlnSerGluAsp 286
 Db 952 GCTAAACAATCGGAAGAC 969

```

; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 68/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12889
; LENGTH: 1807
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; FEATURE:
; LOCATION: 1802, 1803, 1804, 1805, 1806, 1807
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-12889

Alignment Scores:
Pred. No.: 1.47e-119 Length: 1807
Score: 1526.00 Matches: 282
Percent Similarity: 99.30% Conservative: 2
Best Local Similarity: 98.60% Mismatches: 2
Query Match: 98.71% Indels: 0
DB: 47 Gaps: 0

US-10-054-313-1 (1-286) x US-10-198-846-12889 (1-1807)
Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 112 ATGAGCTGGCTTCTGTTCTGCGCCACAGATGCGCTTGGCGCGCTTGGCTGCGCGCGC 171
Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgGlyThrGlyValPhe 40
Db 172 GCGCTCGCGGTTCCGGATGTTCTATCGCTGAGGAGGCGCGCAAGCGCGGCTTTT 231
Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db 232 CTGACCTGGATGAGTGCAGACGACGATGCGCGCTTCTGCTGCGCGATTTAAAGAG 291
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 292 TTTGCCACAGAGGATGAGCGCTTGTGCGGAAATCTGCAAGCGCGGAAAGTTTCA 351
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 352 GAAGGCGCATGAAATCAACATGGAACAGATCGGAGCGGAAAGCGACGACGACTCCGT 411
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 412 GAGCCACTGGATGGAGATGGACATGAAAGCGCAGACGCGGTATGCAAGCACATGAGCGC 471
Qy 121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 472 AGCGTGGAGCGCGCGCCCTCCAGTTAGCAGACACAGCTTTTCTCATGCGGAGACTTCGTC 531
Qy 141 ValValTyrThrAspGlyCysSerSerAsnGlyArgArgGlyArgGlyArgGlyLys 160
Db 532 GTCGCTACATGATGCTGCTGCTCCAGTATGCGCGTATGCGCGCGCGCGCGCGCGATC 591
Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db 592 GCGGTTTACTGGGGCGCAGGCGATCCTTTAAATGTAGGCATATAGACTTCTCTGGCGCGC 651
Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysValSerAlaIleGluGlnAlaLysThrGln 200
Db 652 ACAACACCAAGAGCGGAATTCATGCGCTTGAAGCCATTGAACAAAGCAAGCAACTCAA 711
Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 712 AACATCAATAAAGTGGTCTGTATACAGACAGTATGTTTACGATAAATGGTATTAAC 771
Qy 221 TrpValGlnGlyTrpLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240

```

```

Db 772 TGGTTCAAGGTTGGAGAAATGGTGGAGAACAGTGCAGGAGAAAGAGGTGATCAAC 831
Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db 832 AAAGAGGACTTGTGGCACTGGAGAGGCTTACCAGGGGATGGACATTCAGTGGATGCAT 891
Qy 261 ValProGlyHisSerClyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGly 280
Db 892 GTTCTGGTCACTGGGATTTATAGCAATGAAGAGCTGACAGATTACGAGAGAAGA 951
Qy 281 AlaLysGlnSerGluAsp 286
Db 952 GCTAAACAATCGAAGAC 969

RESULT 6
US-09-770-171-1164
; Sequence 1164, Application US/09770171
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2056-001
; CURRENT APPLICATION NUMBER: US/09/770,171
; CURRENT FILING DATE: 2001-01-29
; PRIOR APPLICATION NUMBER: US 10/178,875
; PRIOR FILING DATE: 2000-01-28
; NUMBER OF SEQ ID NOS: 1568
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1164
; LENGTH: 2124
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-770-171-1164

Alignment Scores:
Pred. No.: 1.77e-119 Length: 2124
Score: 1526.00 Matches: 282
Percent Similarity: 99.30% Conservative: 2
Best Local Similarity: 98.60% Mismatches: 2
Query Match: 98.71% Indels: 0
DB: 33 Gaps: 0

US-10-054-313-1 (1-286) x US-09-770-171-1164 (1-2124)
Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 112 ATGAGCTGGCTTCTGTTCTGCGCCACAGATGCGCTTGGCGCGCTTGGCTGCGCGCGC 171
Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgGlyThrGlyValPhe 40
Db 172 GCGCTCGCGGTTCCGGATGTTCTATCGCTGAGGAGGCGCGCAAGCGCGGCTTTT 231
Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db 232 CTGACCTGGATGAGTGCAGACGACGATGCGCGCTTCTGCTGCGCGATTTAAAGAG 291
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 292 TTTGCCACAGAGGATGAGCGCTTGTGCGGAAATCTGCAAGCGCGGAAAGTTTCA 351
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 352 GAAGGCGCATGAAATCAACATGGAACAGATCGGAGCGGAAAGCGACGACGACTCCGT 411
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 412 GAGCCACTGGATGGAGATGGACATGAAAGCGCAGACGCGGTATGCAAGCACATGAGCGC 471
Qy 121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 472 AGCGTGGAGCGCGCGCCCTCCAGTTAGCAGACACAGCTTTTCTCATGCGGAGACTTCGTC 531
Qy 141 ValValTyrThrAspGlyCysSerSerAsnGlyArgArgGlyArgGlyArgGlyLys 160
Db 532 GTCGCTACATGATGCTGCTGCTCCAGTATGCGCGTATGCGCGCGCGCGCGCGATC 591
Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db 592 GCGGTTTACTGGGGCGCAGGCGATCCTTTAAATGTAGGCATATAGACTTCTCTGGCGCGC 651
Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysValSerAlaIleGluGlnAlaLysThrGln 200
Db 652 ACAACACCAAGAGCGGAATTCATGCGCTTGAAGCCATTGAACAAAGCAAGCAACTCAA 711
Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 712 AACATCAATAAAGTGGTCTGTATACAGACAGTATGTTTACGATAAATGGTATTAAC 771
Qy 221 TrpValGlnGlyTrpLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240

```

```

Db      532  GTCTGCTACCTGATGGCTGCTCCAGTAATGGCTAGAGCCGCGAGCAGGAATC 591
Qy      161  GlyValTyrTrpGlyProGlyHisProLeuValGlyIleArgLeuProGlyArgGln 180
Db      592  GCGCTTACTGGGGCCAGGCCATCTTTAATGTAGGCAATAGACTTCTCTGGGCGGAG 651
Qy      181  ThrAsnGlnArgAlaGluIleHisAlaLaCysLysAlaIleGluGlnAlaLysThrGln 200
Db      652  ACAACCAAGAGCGGAATTCATGCGCTTCAAAAGCAATTAACAAAGCAAGACTCAA 711
Qy      201  AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db      712  AACATCAATAACTGCTTCTGATACAGACATGTTTACGATTAATGGTATTAACCTAAC 771
Qy      221  TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
Db      772  TGGGTTCAAGTTGGAAGAAAATGGGTGGAAGACACAGTGCAGGGAAGAGGTGATCAAC 831
Qy      241  LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db      832  AAAGAGGACTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 891
Qy      261  ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db      892  GTTCTGCTGTCATTCGGGATTTATAGGCATGAAGAGCTGACAGATTAGCCAGAGAGGA 951
Qy      281  AlaLysGlnSerGluAsp 286
Db      952  GCTAAACAATCGGAAGAC 969

```

RESULT 7

```

US-09-785-276A-30281
; Sequence 30281, Application US/09785276A
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCES: MRI-007B
; CURRENT APPLICATION NUMBER: US/09/785,276A
; CURRENT FILING DATE: 2001-02-16
; PRIORITY APPLICATION NUMBER: 60/183,319
; PRIORITY FILING DATE: 2000-02-17
; PRIORITY APPLICATION NUMBER: 60/189,862
; PRIORITY FILING DATE: 2000-03-16
; PRIORITY APPLICATION NUMBER: 60/207,454
; PRIORITY FILING DATE: 2000-05-25
; PRIORITY APPLICATION NUMBER: 60/211,314
; PRIORITY FILING DATE: 2000-06-09
; PRIORITY APPLICATION NUMBER: 60/219,007
; PRIORITY FILING DATE: 2000-07-19
; PRIORITY APPLICATION NUMBER: 60/253,281
; PRIORITY FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30281
; LENGTH: 2129
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2125, 2126, 2127, 2128, 2129
; OTHER INFORMATION: n = A,T,C or G
US-09-785-276A-30281

```

Alignment Scores:

```

Pred. No.: 1,77e-119 Length: 2129
Score: 1526.00 Matches: 282
Percent Similarity: 99.30% Conservative: 2
Best Local Similarity: 98.60% Mismatches: 2

```

```

Query Match: 98.71% Indels: 0
DB: 33 Gaps: 0
US-10-054-313-1 (1-286) x US-09-785-276A-30281 (1-2129)
Qy      1  MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db      112  ATAGGCTGGCTTCTGTCTCTGGCCACAGAGTCGCTTGGCCGCTTGGCTTGGCCGCCG 171
Qy      21  GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
Db      172  GGCTCTCGGGGTTCGGGATGTTCTATGCGGTGAGGAGGGCGCAAGACCGGGGTCTTT 231
Qy      41  LeuThrTrpAsnGlnCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db      232  CTGACCTGGGAATGAGTGCAGAGCACAGGTGGACCGGTTCTCTGCTGCCAGATTTAAGA 291
Qy      61  PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db      292  TTTGCCACAGAGGATGAGGCTGGGCTTTGTTCAGGAATCTGCAAGCCCGGAAGTTTCA 351
Qy      81  GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db      352  GAAAGGCGATGAAATCAACATGGAACAAGATCGAGGCGGAAAGCCAGACGACATCCGT 411
Qy      101  GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db      412  GAGCCACTGGATGGAGATGGACATGAAGCGCAGAGCCGTATGCAAGCACATGAAGCCG 471
Qy      121  SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db      472  AGCGTGGAGCGCGGCTCCAGTTAGCAGAGACACGTTTTTCTTACATGGGAGACTTCGTG 531
Qy      141  ValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
Db      532  GTGCTCTACTGATGGCTGCTCTCCAGTAATGGGCGTAGAAGGCGCGAGCAGGATC 591
Qy      161  GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db      592  GCGCTTACTGGGGCGCGGCTCCCTTTAATGTAGGCAATAGACTTCTCTGGGCGGAG 651
Qy      181  ThrAsnGlnArgAlaGluIleHisAlaCysLysAlaIleGluGlnAlaLysThrGln 200
Db      652  ACAACCAAGAGCGGAATTCATGCGCTTCAAAAGCAATTAACAAAGCAAGACTCAA 711
Qy      201  AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db      712  AACATCAATAACTGCTTCTGATACAGACATGTTTACGATTAATGGTATTAACCTAAC 771
Qy      221  TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
Db      772  TGGGTTCAAGTTGGAAGAAAATGGGTGGAAGACACAGTGCAGGGAAGAGGTGATCAAC 831
Qy      241  LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db      832  AAAGAGGACTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 891
Qy      261  ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db      892  GTTCTGCTGTCATTCGGGATTTATAGGCATGAAGAGCTGACAGATTAGCCAGAGAGGA 951
Qy      281  AlaLysGlnSerGluAsp 286
Db      952  GCTAAACAATCGGAAGAC 969

```

RESULT 8

```

US-10-357-930-30281
; Sequence 30281, Application US/10357930
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR

```


;; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
;; FILE OF INVENTION: HUMAN PROSTATE CANCER
;; FILE REFERENCE: MRI-007BCN
;; CURRENT APPLICATION NUMBER: US/10/357,930
;; CURRENT FILING DATE: 2003-02-04
;; PRIOR APPLICATION NUMBER: 09/785,276
;; PRIOR FILING DATE: 2003-02-16
;; PRIOR APPLICATION NUMBER: 60/183,319
;; PRIOR FILING DATE: 2000-02-17
;; PRIOR APPLICATION NUMBER: 60/189,862
;; PRIOR FILING DATE: 2000-03-16
;; PRIOR APPLICATION NUMBER: 60/207,454
;; PRIOR FILING DATE: 2000-05-25
;; PRIOR APPLICATION NUMBER: 60/211,314
;; PRIOR FILING DATE: 2000-06-09
;; PRIOR APPLICATION NUMBER: 60/219,007
;; PRIOR FILING DATE: 2000-07-18
;; PRIOR APPLICATION NUMBER: 60/285,281
;; PRIOR FILING DATE: 2000-12-13
;; NUMBER OF SEQ ID NOS: 6222
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 30281
;; LENGTH: 2129
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 2125, 2126, 2127, 2128, 2129
;; OTHER INFORMATION: n = A,T,C or G
US-10-357-930-30281

Alignment Scores:
Pred. No.: 1,776-119 Length: 2129
Score: 1526.00 Matches: 282
Percent Similarity: 99.30% Conservative: 2
Best Local Similarity: 98.60% Mismatches: 2
Query Match: 98.71% Indels: 0
DB: 51 Gaps: 0

US-10-054-313-1 (1-286) x US-10-357-930-30281 (1-2129)
QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 112 ATGAGCTGGCTTCTGTTCTGCGCCACAGAGTGGCTTGGCGCGCTTGGCTGGCGCGC 171
QY 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgArgGlyArgGlyValPhe 40
Db 172 GGCTCTCGCGGTTCTGGGATGTTCTATGCCGTGAGGAGGGCGCCAGACCGGGGTCTTT 231
QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60
Db 232 CTGACCTGGAATGAGTGCAGACACAGAGTGGACCGGTTCTCTGCTGCCAGATTAAAGAG 291
QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 292 TTGGCCACAGAGATGAGCGCTGGCGCTTGTGAGGAAATCTGCAAGCCCGGAAAGTTTCA 351
QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 352 GAAGGCGATGAATCAATCATGACACAGATCGAGGGGAAAGCCAGACGCGACTCCGT 411
QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTrpAlaLysHisMetLysPro 120
Db 412 GAGCCACTGGATGGAGATGGACATGAAGAGCGAGCGCGTATGCAAGACACATGAAGCGC 471
QY 121 SerValGluProAlaProValSerArgAspThrPheSerTrpMetGlyAspPheVal 140
Db 472 ACCGTGGAGCGCGCGCTCCAGTTAGCAGACACAGTTTCTCTATCGGAGACTTCGTC 531
QY 141 ValValTrpThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
Db 532 GTCGTCTACACTGATGGCTGCTCCAGTAAATGGCGGTAGAGCGCGCGAGCAAGATC 591

QY 161 GlyValTrpTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db 592 GGGGTTTACTGGGGGCGAGGCCCATCTTTAAATGTAGGCTTAGACTTCTCTGGGCGGAG 651
QY 181 ThrAsnGlnArgAlaGluIleHisAlaCysLysAlaIleGluGlnAlaLysThrGln 200
Db 652 ACAAGCAAGAGAGCGGAAATTCATGAGCGCTGCAAGCCATTGAAACAAGCAAGACTCAA 711
QY 201 AsnIleAsnLysLeuValLeuTrpThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 712 AACATCAATAAATGGTTCTGTATACAGACAGTATGTTACGATAAATGGTATAACTAAC 771
QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
Db 772 TGGGTTCAAGGTTTGGAGAAATATGGTGGAGCAAGTGCAGGAAAGAGAGTGATCAAC 831
QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db 832 AAAGAGGACTTGTGGCACTGGAGAGGCTTACCAGGGGATGGACATTTCAGTGGATGCAT 891
QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db 892 GTTCTGCTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGA 951
QY 281 AlalysGlnSerGluAsp 286
Db 952 GCTAAACAATCGGAAGAC 969

RESULT 9
US-09-763-233-41
; Sequence 41, Application US/09763233
; GENERAL INFORMATION:
; APPLICANT: INCYTE PHARMACEUTICALS, INC.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: CORLEY, Neil J.
; APPLICANT: GUEGLER, Karl J.
; APPLICANT: GORGONE, Gina A.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: SAUGHN, Mariah R.
; APPLICANT: LAL, Preeti
; APPLICANT: BANDMAN, Olga
; APPLICANT: REDDY, Roopa
; APPLICANT: AZIMZAI, Valda
; APPLICANT: SHIH, Leo L.
; APPLICANT: YANG, Junming
; TITLE OF INVENTION: HUMAN RNA-ASSOCIATED PROTEINS
; FILE REFERENCES: PF-0579 PCT
; CURRENT APPLICATION NUMBER: US/09/763,233
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 60/097,550; 60/115,639
; PRIOR FILING DATE: 1998-08-21; 1999-01-12
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PERL Program
; SEQ ID NO 41
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte Identification No.: 2073417CBI
US-09-763-233-41

Alignment Scores:
Pred. No.: 1,596-119 Length: 1150
Score: 1523.00 Matches: 281
Percent Similarity: 99.30% Conservative: 3
Best Local Similarity: 98.25% Mismatches: 2
Query Match: 98.51% Indels: 0
DB: 33 Gaps: 0

US-10-054-313-1 (1-286) x US-09-763-233-41 (1-1150)

```
Qy 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 85 ATGAGCTGGCTTCTGCTGGCCACAGAGTGGCTTGGCCGCTTGGCCGCGCCG 144
Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgValPhe 40
Db 145 GGCCTCTCGGGGTTCTGCGGATGTTCTATCGCGTGGAGGCGCGCAGACCGGGTCTTT 204
Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60
Db 205 CTGACCTGGATGAGTGCAGAGCACAGGTGGACCGGTTCTCTGCTGCCAGATTAAAG 264
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 265 TTTGCCACAGAGGATGAGGCTGGGCTTGTCTAGGAAATCTGCAAGCCCGGAAGTTTCA 324
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 325 GAAGGGCATGAAATCAACATGGACAGAAATCGAGGCGGAAAGCCAGACGACCTCCGT 384
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 385 GAGCCACTGGATGGAGATGACATGAAAGCCGAGAGCGGTATGCAAGCCGAGCAATGAAGCCG 444
Qy 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 445 AGCATGGAGCCGGCGCTCCATGTAGCAGACAGCTTTCTCTACATGGAGACTTCGTC 504
Qy 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160
Db 505 GTGCTCTACACTGATGAGTGGCTTGTCTAGGAAATCTGCAAGCCCGGAAGTTTCA 564
Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLysArgLeuProGlyArgGln 180
Db 565 GGCCTTTACTGGGGCCAGCGCATCTTTAAATGTAGGCATTTAGACTTCTCTGGCGGCGAG 624
Qy 181 ThrAsnGlnArgAlaGluLeuHisAlaLysLysAlaLysAlaLysThrGln 200
Db 625 ACAACCAAGAGCGGAAATTCATGAGCCTGCAAGCCATTGAACCAAGCAAGACTCAA 684
Qy 201 AsnLeuAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 685 AACATCAATAAAGTGGTCTCTGTATACAGACAGTATGTTTACGATAAATGGTATTAAC 744
Qy 221 TrpValGlnGlyTyrLysLysAsnGlyTyrLysThrSerAlaGlyLysGluValIleAsn 240
Db 745 TGGGTTCAAGTTTGAAGAAATGGTGGAGACAAGTGGAGGAAAGAGGTGATCAAC 804
Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db 805 AAAGAGGACTTTTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTGAGTGGATGCAT 864
Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db 865 GTTCTCTGGTCACTCGGATTTATAGCAATGAGAGAGCTGACAGATTATCCAGAGAGGA 924
Qy 281 AlaLysGlnSerGluAsp 286
Db 925 GCTAAACAATCGGAAGAC 942
```

RESULT 10

US-60-115-639-32
; Sequence 32, Application US/60115639
; GENERAL INFORMATION:
; APPLICANT: Preeti Lal
; APPLICANT: Henry Yue
; APPLICANT: Olga Bandman
; APPLICANT: Y. Tom Tang
; APPLICANT: Jennifer L. Hillman
; APPLICANT: Roopa Reddy
; APPLICANT: Mariah R. Baughn

APPLICANT: Yalda Azimzai
APPLICANT: Leo Shih
APPLICANT: Junming Yang
APPLICANT: Aina Lu
TITLE OF INVENTION: HUMAN RNA-ASSOCIATED PROTEINS
FILE REFERENCE: PF-0655 P
CURRENT APPLICATION NUMBER: US/60/115,639
CURRENT FILING DATE: 1999-01-12
NUMBER OF SEQ ID NOS: 40
SOFTWARE: PERL Program
SEQ ID NO 32
LENGTH: 1150
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE: -
OTHER INFORMATION: 2073417
US-60-115-639-32

Alignment Scores:
Pred. No.: 1,59e-119 Length: 1150
Score: 1523.00 Matches: 281
Percent Similarity: 99.30% Conservative: 3
Best Local Similarity: 98.25% Mismatches: 2
Query Match: 98.51% Indels: 0
DB: 65 Gaps: 0

US-10-054-313-1 (1-286) x US-60-115-639-32 (1-1150)

```
Qy 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 85 ATGAGCTGGCTTCTGCTGGCCACAGAGTGGCTTGGCCGCTTGGCCGCGCCG 144
Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgValPhe 40
Db 145 GGCCTCTCGGGGTTCTGCGGATGTTCTATCGCGTGGAGGCGCGCAGACCGGGTCTTT 204
Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60
Db 205 CTGACCTGGATGAGTGCAGAGCACAGGTGGACCGGTTCTCTGCTGCCAGATTAAAG 264
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 265 TTTGCCACAGAGGATGAGGCTGGGCTTGTCTAGGAAATCTGCAAGCCCGGAAGTTTCA 324
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 325 GAAGGGCATGAAATCAACATGGACAGAAATCGAGGCGGAAAGCCAGACGACCTCCGT 384
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 385 GAGCCACTGGATGGAGATGACATGAAAGCCGAGAGCGGTATGCAAGCCGAGCAATGAAGCCG 444
Qy 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 445 AGCATGGAGCCGGCGCTCCATGTAGCAGACAGCTTTCTCTACATGGAGACTTCGTC 504
Qy 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160
Db 505 GTGCTCTACACTGATGAGTGGCTTGTCTAGGAAATCTGCAAGCCCGGAAGTTTCA 564
Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLysArgLeuProGlyArgGln 180
Db 565 GGCCTTTACTGGGGCCAGCGCATCTTTAAATGTAGGCATTTAGACTTCTCTGGCGGCGAG 624
Qy 181 ThrAsnGlnArgAlaGluLeuHisAlaLysLysAlaLysAlaLysThrGln 200
Db 625 ACAACCAAGAGCGGAAATTCATGAGCCTGCAAGCCATTGAACCAAGCAAGACTCAA 684
Qy 201 AsnLeuAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 685 AACATCAATAAAGTGGTCTCTGTATACAGACAGTATGTTTACGATAAATGGTATTAAC 744
Qy 221 TrpValGlnGlyTyrLysLysAsnGlyTyrLysThrSerAlaGlyLysGluValIleAsn 240
```

Db 745 TGGGTTCAAGTTTGAAGAAATGGGTGGAAGCAAGTGCAGGAAAGAGTGTATCAAC 804
Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db 805 AAAGAGACATTGTGGGACCTGGAGAGGCTTACCAGGGGATGGACATTCACTGATGCAT 864
Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db 865 GTTCCTGTCTATTCGGGATTTATAGCAATTAAGCAAGTGCACAGATTAGCCAGAGAAGA 924
Qy 281 AlAlysGlnSerGluAsp 286
Db 925 GCTAAACATCGGAGAC 942

RESULT 11

PCT-US02-31373-85

Sequence 85, Application PC/TUS0231373

GENERAL INFORMATION:

APPLICANT: Curagen Corporation

APPLICANT: Smithson, Glennda

APPLICANT: Millet, Isabelle

APPLICANT: Peyman, John A.

APPLICANT: Kekuda, Ramesh

APPLICANT: Ju, Jingfang

APPLICANT: Li, Li

APPLICANT: Guo, Xiaojia (Sasha)

APPLICANT: Patturajan, Meera

APPLICANT: Spytek, Kimberly A.

APPLICANT: Edinger, Shlomit R.

APPLICANT: Ellerman, Karen

APPLICANT: Malyankar, Uriel M.

APPLICANT: Ort, Tatiana

APPLICANT: Gorman, Linda

APPLICANT: Zernusen, Bryan D.

APPLICANT: Anderson, David W.

APPLICANT: Zhong, Mei

APPLICANT: Catterton, Elina

APPLICANT: Ji, Weizhen

APPLICANT: Miller, Charles E.

APPLICANT: Rastelli, Luca

APPLICANT: Stone, David J.

APPLICANT: Pena, Carol E. A.

APPLICANT: Shenoy, Suresh G.

APPLICANT: Shimkets, Richard A.

APPLICANT: Rothenberg, Mark E.

APPLICANT: Leach, Martin D.

APPLICANT: Agee, Michele L.

APPLICANT: Berghs, Constance

TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

FILE REFERENCE: 21402-462C-061

CURRENT APPLICATION NUMBER: PCT/US02/31373

CURRENT FILING DATE: 2003-09-28

PRIOR APPLICATION NUMBER: 10/262,511

PRIOR FILING DATE: 2002-10-01

PRIOR APPLICATION NUMBER: 60/326,483

PRIOR FILING DATE: 2001-10-02

PRIOR APPLICATION NUMBER: 60/373,815

PRIOR FILING DATE: 2002-04-19

PRIOR APPLICATION NUMBER: 60/327,917

PRIOR FILING DATE: 2001-10-09

PRIOR APPLICATION NUMBER: 60/381,642

PRIOR FILING DATE: 2002-05-17

PRIOR APPLICATION NUMBER: 60/328,029

PRIOR FILING DATE: 2002-10-09

PRIOR APPLICATION NUMBER: 60/381,038

PRIOR FILING DATE: 2002-05-16

PRIOR APPLICATION NUMBER: 60/328,056

PRIOR FILING DATE: 2001-10-09

PRIOR APPLICATION NUMBER: 60/373,260

PRIOR FILING DATE: 2002-04-17

PRIOR APPLICATION NUMBER: 60/373,826

PRIOR FILING DATE: 2002-04-19

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 439

; SOFTWARE: CuraSeqList version 0.1

; SEQ ID NO 85

; LENGTH: 965

; TYPE: DNA

; ORGANISM: CG144997-02

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (10)..(867)

PCT-US02-31373-85

Alignment Scores:

Pred. No.: 1.58e-119 Length: 965

Score: 1522.00 Matches: 281

Percent Similarity: 98.95% Conservative: 2

Best Local Similarity: 98.25% Mismatches: 3

Query Match: 98.45% Indels: 0

DB: 1 Gaps: 0

US-10-054-313-1 (1-286) x PCT-US02-31373-85 (1-965)

Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20

Db 10 ATGAGCTGTTTCTGTCTCTGCCACACAGTGGCTTGGCGGCTTGCCTGCGCGCGC 69

Qy 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgGlyArgGlyThrGlyValPhe 40

Db 70 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGAGGGCGCGAAGACCGGGGTCTTT 129

Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60

Db 130 CTGACCTGCAATGAGTGCAGACACAGGTGGACCGGTTTCTGTCTGCCAGATTTAAGAAG 189

Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80

Db 190 TTTGCCACAGAGATGAGGCTTGGGCTTGTTCAGGAAATCTGCAAGCCCGGAAGTTTCA 249

Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100

Db 250 GAAGGCGATGAAATCAACATGCACAGAAATCGGAGGCGAAGCCAGCAGGACTCGT 309

Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120

Db 310 GAGCCACTGGATGGAGATGGACATGAAGCGCAGAGCCGTATGCAAGACACATGAAGCGC 369

Qy 121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140

Db 370 AGCGTGAGCGCGGCTCCAGTTAGCAGACACGTTTTCTTACATGGGAGACTTCGTC 429

Qy 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160

Db 430 GTGCTCTACACTGATGGTGTCTCCAGTAATGGCGGTAGAAGGCGCGGAGCAAGATC 489

Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180

Db 490 GCGGTTTACTGGGCGCGGCGCATCTTTAAATAGGACATTAGACTTCTCTGGCGGCGAG 549

Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200

Db 550 ACAACCAAGAGCGGAATTCATGCAGCTGCAGAGCCATTGAACAAGAAAGACTCAA 609

Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220

Db 610 AACATCAATAAACTGGTCTCTGTATACAGACAGTATGTTTACGATAAATGGTATACTAC 669

Qy 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240

Db 670 TGGGTTCAAGTTGGAGAAAATGGGTGGAGACAGTGCAGGGAAGAGGTGATCAAC 729

Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260

Db 730 AAAGAGGACTTTGTGCACTGGAGAGGCTTACCAGGGATGGACATTCAGTGGATGCAT 789

QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGlyGly 280
 Db 790 GTTCCTGGTCACTTCGGGATTTATAGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA 849

QY 281 AlaLysGlnSerGluAsp 286
 Db 850 GCTAAACAATCGAAGAC 867

RESULT 12

US-10-262-511-85

Sequence 85, Application US/10262511

GENERAL INFORMATION:

APPLICANT: Smithson, Glennda

APPLICANT: Millet, Isabelle

APPLICANT: Peyman, John A.

APPLICANT: Kekuda, Ramesh

APPLICANT: Ju, Jingfang

APPLICANT: Li, Li

APPLICANT: Guo, Xiaojia (Sasha)

APPLICANT: Patturajan, Meera

APPLICANT: Spytek, Kimberly A.

APPLICANT: Edinger, Shlomit R.

APPLICANT: Ellerman, Karen

APPLICANT: Malvankar, Uriel M.

APPLICANT: Ort, Tatiana

APPLICANT: Gorman, Linda

APPLICANT: Zernhusen, Bryan D.

APPLICANT: Anderson, David W.

APPLICANT: Zhong, Mei

APPLICANT: Catterton, Elina

APPLICANT: Ji, Weizhen

APPLICANT: Miller, Charles E.

APPLICANT: Rastelli, Luca

APPLICANT: Stone, David J.

APPLICANT: Pena, Carol E. A.

APPLICANT: Shency, Suresh G.

APPLICANT: Shimkets, Richard A.

APPLICANT: Rothenberg, Mark E.

APPLICANT: Leach, Martin D.

APPLICANT: Agee, Michele L.

APPLICANT: Bergha, Constance

TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

FILE REFERENCE: 21402-462C

CURRENT APPLICATION NUMBER: US/10/262,511

CURRENT FILING DATE: 2003-05-28

PRIOR APPLICATION NUMBER: 60/326,483

PRIOR FILING DATE: 2001-10-02

PRIOR APPLICATION NUMBER: 60/373,815

PRIOR FILING DATE: 2002-04-19

PRIOR APPLICATION NUMBER: 60/327,917

PRIOR FILING DATE: 2001-10-09

PRIOR APPLICATION NUMBER: 60/381,642

PRIOR FILING DATE: 2002-05-17

PRIOR APPLICATION NUMBER: 60/328,029

PRIOR FILING DATE: 2002-10-09

PRIOR APPLICATION NUMBER: 60/381,038

PRIOR FILING DATE: 2002-05-16

PRIOR APPLICATION NUMBER: 60/328,056

PRIOR FILING DATE: 2001-10-09

PRIOR APPLICATION NUMBER: 60/373,260

PRIOR FILING DATE: 2002-04-17

PRIOR APPLICATION NUMBER: 60/373,826

PRIOR FILING DATE: 2002-04-19

PRIOR APPLICATION NUMBER: 60/327,435

PRIOR FILING DATE: 2001-10-05

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 439

SOFTWARE: Curaseq version 0.1

SEQ ID NO 85

LENGTH: 965

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (10)..(867)
 US-10-262-511-85

Alignment Scores:

Pred. No.: 1,586-119 Length: 965
 Score: 1522.00 Matches: 281
 Percent Similarity: 98.95% Conservative: 2
 Best Local Similarity: 98.25% Mismatches: 3
 Query Match: 98.45% Indels: 0
 DB: 49 Gaps: 0

US-10-054-313-1 (1-286) x US-10-262-511-85 (1-965)

| | | | |
|----|-----|--|-----|
| QY | 1 | MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgGly | 20 |
| Db | 10 | ATGAGCTGGTTCTGTCTGCGCCACAGAGTCGCTTGGCCGCTTGGCCCTCCGCCCGC | 69 |
| QY | 21 | GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgGlyThrGlyValPhe | 40 |
| Db | 70 | GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGCGCAAGACCGGGGCTTT | 129 |
| QY | 41 | LeuThrTrpAsnGlyCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys | 60 |
| Db | 130 | CTGACCTGGAATGAGTGCAGACACAGGTGGACCGGTTTCTCTCTCCAGATTAAAGAAG | 189 |
| QY | 61 | PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer | 80 |
| Db | 190 | TTTGCCACAGAGATGAGGCTGGGCTTTGTTCAGGAATCTGCAAGCCCGGAATTTC | 249 |
| QY | 81 | GluGlyHisGluAsnGlnHisGlyGlnGlnSerGluAlaLysProGlyLysArgLeuArg | 100 |
| Db | 250 | GAGGGCATGAAAATCAACATGACAGATCGAGGCGGAAGCCAGCAAGCGACTCCGT | 309 |
| QY | 101 | GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro | 120 |
| Db | 310 | GAGCCACTGGATGAGATGAGCATGAAGCGCAGAGCCGATGCAAGACACATGAAGCCG | 369 |
| QY | 121 | SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal | 140 |
| Db | 370 | AGCGTGAGCGCGCCCTCCAGTTAGCAGAGACACAGTTTCTTACATGGGAGACTTCGTC | 429 |
| QY | 141 | ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyLe | 160 |
| Db | 430 | GTCGCTCTACACTGATGGCTGCTCTCCAGTAATGGCGGTAGAGGCGCGGAGCAGGAATC | 489 |
| QY | 161 | GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln | 180 |
| Db | 490 | GCGCTTTACTGGGGCCAGGCCATCTCTTTAAATGTAGGCATTAGACTTCTGGGCGCAG | 549 |
| QY | 181 | ThrAsnGlnArgAlaGluIleHisAlaCysLysAlaIleGluGlnAlaLysThrGln | 200 |
| Db | 550 | ACAAACCAAGAGCGGAATTCATGCGAGCTGCAAGCCATTGAACCAAGCAAGACTCAA | 609 |
| QY | 201 | AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn | 220 |
| Db | 610 | AACATCAATAAATCGTTCTGTATACAGACAGTATGTTTACGATAAATGGTAACTAAC | 669 |
| QY | 221 | TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn | 240 |
| Db | 670 | TGGGTTCAAGGTTGGAAGAAAATGGGTGGAAGCAAGAGTCAGGGAAGAGGTGATCAAC | 729 |
| QY | 241 | LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis | 260 |
| Db | 730 | AAAGAGACTTTGTGGCAGCTGGAGAGCTTACCAGGGGATGGACATTCAGTGGATGCAT | 789 |
| QY | 261 | ValProGlyHisSerGlyPheIleGlyAsnGlnGluAlaAspArgLeuAlaArgGlyGly | 280 |
| Db | 790 | GTTCCTGGTCACTTCGGGATTTATAGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA | 849 |
| QY | 281 | AlaLysGlnSerGluAsp 286 | |

Db 850 GCTAAACAATCGGAAGAC 867

RESULT 13

US-10-262-511A-85

; Sequence 85, Application US/10262511A

; GENERAL INFORMATION:

; APPLICANT: Smithson, Glenda

; APPLICANT: Millet, Isabelle

; APPLICANT: Peyman, John A.

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Ju, Jingfang

; APPLICANT: Li, Li

; APPLICANT: Guo, Xiaojia (Sasha)

; APPLICANT: Patturajan, Meera

; APPLICANT: Spvtek, Kimberly A.

; APPLICANT: Edinger, Shlomit R.

; APPLICANT: Ellerman, Karen

; APPLICANT: Malyankar, Uriel M.

; APPLICANT: Ort, Tatiana

; APPLICANT: Gorman, Linda

; APPLICANT: Zerhusen, Bryan D.

; APPLICANT: Anderson, David W.

; APPLICANT: Zhong, Mei

; APPLICANT: Catterton, Elina

; APPLICANT: Ji, Weizhen

; APPLICANT: Miller, Charles E.

; APPLICANT: Rastelli, Luca

; APPLICANT: Stone, David J.

; APPLICANT: Pena, Carol E. A.

; APPLICANT: Shenov, Suresh G.

; APPLICANT: Shinkets, Richard A.

; APPLICANT: Rothenberg, Mark E.

; APPLICANT: Leach, Martin D.

; APPLICANT: Agee, Michele L.

; APPLICANT: Berghs, Constance

; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

; FILE REFERENCE: 21402-462C

; CURRENT APPLICATION NUMBER: US/10/262,511A

; CURRENT FILING DATE: 2002-10-01

; PRIOR APPLICATION NUMBER: 60/326,483

; PRIOR FILING DATE: 2001-10-02

; PRIOR APPLICATION NUMBER: 60/373,815

; PRIOR FILING DATE: 2002-04-19

; PRIOR APPLICATION NUMBER: 60/327,917

; PRIOR FILING DATE: 2001-10-09

; PRIOR APPLICATION NUMBER: 60/381,642

; PRIOR FILING DATE: 2002-05-17

; PRIOR APPLICATION NUMBER: 60/328,029

; PRIOR FILING DATE: 2002-10-09

; PRIOR APPLICATION NUMBER: 60/381,038

; PRIOR FILING DATE: 2002-05-16

; PRIOR APPLICATION NUMBER: 60/328,056

; PRIOR FILING DATE: 2001-10-09

; PRIOR APPLICATION NUMBER: 60/373,260

; PRIOR FILING DATE: 2002-04-17

; PRIOR APPLICATION NUMBER: 60/324,826

; PRIOR FILING DATE: 2002-04-19

; PRIOR APPLICATION NUMBER: 60/327,435

; PRIOR FILING DATE: 2001-10-05

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 439

; SOFTWARE: Curaseq1ist version 0.1

; SEQ ID NO 85

; LENGTH: 965

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (10)..(867)

US-10-262-511A-85

Alignment Scores:

Pred. No.: 1.58e-119 Length: 965

Score: 1522.00 Matches: 281

Percent Similarity: 98.95% Conservativity: 2

Best Local Similarity: 98.25% Mismatches: 3

Query Match: 49 Indels: 0

DB: Gaps: 0

US-10-054-313-1 (1-286) x US-10-262-511A-85 (1-965)

QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20

DB 10 ATGAGCTGGTTCTGTTCTGCGCCACAGATCGCTTGGCGCGCTTCCCTTCCCGCCGC 69

QY 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgArgGlyArgLysThrGlyValPhe 40

DB 70 GGCTCTCGGGGTTCCGGATGTTCTATCCCTGAGAGGGGCGGCACAGACCGGGGTCCTT 129

QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAlaAspArgPheProAlaAlaArgPheLys 60

DB 130 CTGACCTGGATGATGTCGAGAGCACAGGTGGACCGGTTCTCTGCGCCAGATTTAAGAAG 189

QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80

DB 190 TTTGCCACAGAGGATGAGGCTGGCCCTTTGTAGAAATCTGCAAGCCCGAACTTCA 249

QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100

DB 250 GAAGGCGATGAAATCAACATGGACAGAAATCGGAGGCGAAAGCCAGACGACTCCGT 309

QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTrpAlaLysHisMetLysPro 120

DB 310 GAGCCACTGGATGGAGATGGACATGAAGCCGACAGAGCCGTATGCAAGCACATGAGCG 369

QY 121 SerValGluProAlaProValSerArgAspThrPheSerTrpMetGlyAspPheVal 140

DB 370 AGCGTGGAGCGCGCCCTCCAGTTAGCAGAGACACAGTTTCTTCTACATGGGAGACTTCGTC 429

QY 141 ValValTrpThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyLys 160

DB 430 GTCTGTCTACATGATGGTGTCTGCTCCAGTAAATGGGCGTAGAGCCCGGAGCAGGAATC 489

QY 161 GlyValTrpTrpGlyProGlyHisProLeuAsnValGlyLysArgLeuProGlyArgGln 180

DB 490 GCGGTTTACTGGGGCCAGGCGCATCTTTAAATGTAGGCATTTAGACTTCTTGGGCGGAG 549

QY 181 ThrAsnGlnArgAlaGluLeuHisAlaAlaCysLysAlaLysGluAlaLysThrGln 200

DB 550 ACAACCAAGAGCGGAAATTCATGCGCCTGCAAGCCATTGAACAAAGAAAGACTCAA 609

QY 201 AsnIleAsnLysLeuValLeuTrpThrAspSerMetPheThrIleAsnGlyLysThrAsn 220

DB 610 AACATCAATAAACTGGTTCTGTATACAGACAGTAGTATGTTTACGATAATAGTATAACTAAC 669

QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValLysAsn 240

DB 670 TGGGTTCAAGGTGGAGAAATAATGGTGGAGAACACAGTCGAGGAAAGAGGTGATCAAC 729

QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260

DB 730 AAAGAGACTTTGTGCGACTGGAGAGGCTTACCCAGGCGGATGGACATTCAGTGGATGCAT 789

QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280

DB 790 GTTCTGTGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGA 849

QY 281 AlaLysGlnSerGluAsp 286

DB 850 GCTAAACAATCGGAAGAC 867

RESULT 14

PCT-US02-18947-841

; Sequence 841, Application PC/TUS0218947

; GENERAL INFORMATION:

; APPLICANT: Rosetta Inpharmatics

;; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients

;; FILE REFERENCE: 9301-175-228
;; CURRENT APPLICATION NUMBER: PCT/US02/18947
;; CURRENT FILING DATE: 2002-06-14
;; PRIOR APPLICATION NUMBER: 603,800,770
;; PRIOR FILING DATE: 2002-05-14
;; NUMBER OF SEQ ID NOS: 2699
;; SEQ ID NO 841
;; LENGTH: 1147
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; PUBLICATION INFORMATION:
;; DATABASE ACCESSION NUMBER: NM_002936
;; DATABASE ENTRY DATE: 2001-06-18
PCT-US02-18947-841

Alignment Scores:
Pred. No.: 1,93e-119 Length: 1147
Score: 1522.00 Matches: 281
Percent Similarity: 98.95% Conservative: 2
Best Local Similarity: 98.25% Mismatches: 3
Query Match: 98.45% Indels: 0
DB: 1 Gaps: 0

US-10-054-313-1 (1-286) x PCT-US02-18947-841 (1-1147)

QY 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
DB 62 ATGAGCTGGTTCTGTTCTGTCGCCACAGAGTCGCTTGGCCGCTTCCCTTGGCCGCGC 121
QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
DB 122 GGCCTCTGGGGTTCGGGATGTTCTATGCCGTGAGGAGGCGCGCAAGACCGGGTCTTT 181
QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60
DB 182 CTGACCTGGAATGAGTGCAGACACAGGTGGACCGGTTCTGCTGCCAGATTTAAGAAG 241
QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
DB 242 TTGGCCACAGAGATGAGCGCTTGGGCTTTGTCAGGAATCTGCAAGCCCGGAGTTTCA 301
QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysPheArgLys 100
DB 302 GAAGGCGATGAAATCAACATGACAAAGATCGGAGGCGCGCGCAAGACCGGGTCTTT 181
QY 101 GluProLeuAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
DB 362 GAGCCACTGATGGATGGACATGATGAAGCGGAGCGCGTATGCAAGCACATGAGCGG 421
QY 121 SerValGluProAlaProValSerArgPheThrPheSerTyrMetGlyAspPheVal 140
DB 422 AGCGTGGAGCGCGCGCTCCAGTTAGCAGACACCGTTTCTTACATGGAGACTTCGTC 481
QY 141 ValValTyrThrAspGlyCysSerSerAsnGlyArgArgProAlaGlyLys 160
DB 482 GTGCTCTACACTGATGCTCTCTCCATGATATGGCGGTAGAGGCGCGCGACAGGATC 541
QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLysArgLeuProGlyArgGln 180
DB 542 GGCCTTTACTGGGGCGAGCCCATCTTTAAATGTAGGCATTAGACTTCTGGGCGGCG 601
QY 181 ThrAsnGlnArgAlaGluLysHisAlaAlaCysLysAlaLysGluAlaLysThrGln 200
DB 602 ACAACCAAGAGCGGAAATTCATGCGACGCTCGAAAGCCATTGAAACAAGAAAGACTCAA 661
QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyLysThrAsn 220
DB 662 AACATCAATAAATGCTTCTGTATACACAGCATGATGTTTACGATAAATGGTATACTAAC 721
QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValLysAsn 240
DB 722 TGGGTTTCAGGTTTGAAGAAAAATGGGTGGAAGACAAGTGCAGGAAAGAGGTGATCAAC 781

QY 241 LysGluAspPheValAlaLeuGluArgLysThrGlnGlyMetAspIleGlnTrpMetHis 260
DB 782 AAAGAGGACTTTGTCGCACTCGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 841
QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGly 280
DB 842 GTTCCTGGTTCATTCGGGATTTATAGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGA 901
QY 281 AlaLysGlnSerGluAsp 286
DB 902 GCTAAACAATCGAGAGAC 919

RESULT 15

US-10-172-118-841
;; Sequence 841, Application US/10172118
;; GENERAL INFORMATION:
;; APPLICANT: Dai, Hongyue
;; APPLICANT: He, Yudong
;; APPLICANT: Linsley, Peter
;; APPLICANT: Mao, Mao
;; APPLICANT: Roberts, Chris
;; APPLICANT: Van 't Veer, Laura
;; APPLICANT: Van de Vijver, Marc
;; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
;; FILE REFERENCE: 9301-175-999
;; CURRENT APPLICATION NUMBER: US/10/172,118
;; CURRENT FILING DATE: 2002-06-14
;; PRIOR APPLICATION NUMBER: 60/380,770
;; PRIOR FILING DATE: 2002-05-14
;; NUMBER OF SEQ ID NOS: 2699
;; SEQ ID NO 841
;; LENGTH: 1147
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; PUBLICATION INFORMATION:
;; DATABASE ACCESSION NUMBER: NM_002936
;; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-841

Alignment Scores:
Pred. No.: 1,93e-119 Length: 1147
Score: 1522.00 Matches: 281
Percent Similarity: 98.95% Conservative: 2
Best Local Similarity: 98.25% Mismatches: 3
Query Match: 98.45% Indels: 0
DB: 1 Gaps: 0

US-10-054-313-1 (1-286) x US-10-172-118-841 (1-1147)

QY 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
DB 62 ATGAGCTGGTTCTGTTCTGTCGCCACAGAGTCGCTTGGCCGCTTCCCTTGGCCGCGC 121
QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
DB 122 GGCCTCTGGGGTTCGGGATGTTCTATGCCGTGAGGAGGCGCGCAAGACCGGGTCTTT 181
QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60
DB 182 CTGACCTGGAATGAGTGCAGACACAGGTGGACCGGTTCTGCTGCCAGATTTAAGAAG 241
QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
DB 242 TTGGCCACAGAGATGAGCGCTTGGGCTTTGTCAGGAATCTGCAAGCCCGGAGTTTCA 301
QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysPheArgLys 100
DB 302 GAAGGCGATGAAATCAACATGACAAAGATCGGAGGCGCGCGCAAGACCGGGTCTCG 361
QY 101 GluProLeuAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
DB 362 GAGCCACTGATGGATGGACATGATGAAGCGGAGCGCGTATGCAAGCACATGAGCGG 421

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 17, 2003, 18:09:14 ; Search time 2512 Seconds
(without alignments)
2767.150 Million cell updates/sec

Title: US-10-054-313-1

Perfect score: 1546

Sequence: 1 MSWLLFLAHRVALAALPCRR.....FIGNEADRLAREGAKQSED 286

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh

-MODEL=frame+pn.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US10054313/runat 17122003 150744 24191/app query.fasta_1.455
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054313 @C@N 1 1 2810 @runat 17122003 150744 24191 -NCPU=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba.*
2: em_esthum.*
3: em_estcin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_est3.*
12: gb_est4.*
13: gb_est5.*
14: gb_est6.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_pbg.*
27: em_gss_vrl.*
28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|--------|-------|--------|----|----------|--------------------|
| 1 | 1501 | 97.1 | 1201 | 13 | BX394121 | BX394121 BX394121 |
| 2 | 1498 | 96.9 | 1201 | 9 | AL560874 | AL560874 AL560874 |
| 3 | 1458 | 94.1 | 1054 | 12 | BM541420 | BM541420 AGENCOURT |
| 4 | 1446 | 93.5 | 1060 | 12 | BM810651 | BM810651 AGENCOURT |
| 5 | 1432 | 92.6 | 952 | 9 | AL554334 | AL554334 AGENCOURT |
| 6 | 1415.5 | 91.6 | 944 | 12 | BG282086 | BG282086 602752428 |
| 7 | 1406 | 90.9 | 1196 | 13 | BX416456 | BX416456 BX416456 |
| 8 | 1345 | 87.0 | 893 | 13 | BUI75613 | BUI75613 AGENCOURT |
| 9 | 1340 | 86.7 | 949 | 12 | BG831321 | BG831321 602766160 |
| 10 | 1288 | 83.3 | 765 | 14 | CA775004 | CA775004 ip11g06.y |
| 11 | 1281 | 82.9 | 935 | 13 | BQ21568 | BQ21568 AGENCOURT |
| 12 | 1239.5 | 80.2 | 779 | 10 | BE746490 | BE746490 601579445 |
| 13 | 1229 | 79.5 | 818 | 13 | BU595053 | BU595053 AGENCOURT |
| 14 | 1222 | 79.0 | 916 | 12 | BM451915 | BM451915 AGENCOURT |
| 15 | 1215 | 78.6 | 869 | 10 | BF984400 | BF984400 602307923 |
| 16 | 1210 | 78.3 | 707 | 12 | BI223765 | BI223765 602943149 |
| 17 | 1200 | 77.6 | 959 | 10 | B5689462 | B5689462 602186837 |
| 18 | 1198.5 | 77.5 | 714 | 10 | BE778327 | BE778327 601463720 |
| 19 | 1195 | 77.3 | 946 | 10 | BF038506 | BF038506 601460421 |
| 20 | 1155 | 74.7 | 764 | 12 | BG767598 | BG767598 602741757 |
| 21 | 1153 | 74.6 | 1198 | 12 | BM805842 | BM805842 AGENCOURT |
| 22 | 1110.5 | 71.8 | 693 | 12 | BG768406 | BG768406 602742157 |
| 23 | 1100 | 71.2 | 864 | 10 | BG420594 | BG420594 602448473 |
| 24 | 1090.5 | 70.5 | 782 | 13 | BU329741 | BU329741 AGENCOURT |
| 25 | 1089.5 | 70.5 | 772 | 14 | CB961021 | CB961021 AGENCOURT |
| 26 | 1083.5 | 70.1 | 792 | 10 | BE407918 | BE407918 601292660 |
| 27 | 1067 | 69.0 | 613 | 12 | BM782115 | BM782115 K-EST0058 |
| 28 | 1058 | 68.4 | 652 | 12 | BI518784 | BI518784 603061972 |
| 29 | 1058 | 68.4 | 656 | 12 | BG824728 | BG824728 602728713 |
| 30 | 1052.5 | 68.1 | 937 | 10 | BE786259 | BE786259 601474394 |
| 31 | 1040.5 | 67.3 | 704 | 10 | BF129039 | BF129039 601811367 |
| 32 | 1018 | 65.8 | 1209 | 12 | BM454170 | BM454170 AGENCOURT |
| 33 | 1003.5 | 64.9 | 868 | 13 | BQ942688 | BQ942688 AGENCOURT |
| 34 | 1002 | 64.8 | 783 | 14 | CB529180 | CB529180 UI-H-F72 |
| 35 | 996 | 64.4 | 1364 | 11 | AK011680 | AK011680 Mus muscu |
| 36 | 992 | 64.2 | 927 | 12 | BI558492 | BI558492 603240375 |
| 37 | 973 | 62.9 | 823 | 13 | BQ896738 | BQ896738 AGENCOURT |
| 38 | 968.5 | 62.6 | 1102 | 12 | BI410112 | BI410112 602964039 |
| 39 | 967 | 62.5 | 619 | 10 | BG719686 | BG719686 60288886 |
| 40 | 960.5 | 62.1 | 718 | 12 | BG778946 | BG778946 602867481 |
| 41 | 952 | 61.6 | 801 | 10 | BF690400 | BF690400 602186827 |
| 42 | 949 | 61.4 | 534 | 12 | BM839989 | BM839989 K-EST0116 |
| 43 | 945 | 61.1 | 807 | 9 | AW006811 | AW006811 wt07f11.x |
| 44 | 944 | 61.1 | 775 | 14 | CA435275 | CA435275 UI-H-DP0 |
| 45 | 942 | 60.9 | 574 | 10 | AW965779 | AW965779 EST377852 |

ALIGNMENTS

RESULT 1
BX394121 BX394121 1201 bp mRNA linear EST 13-MAY-2003
LOCUS BX394121 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
DEFINITION CDNA clone CS06C010Y123 5-PRIME, mRNA sequence.
ACCESSION BX394121
VERSION BX394121.1 GI:30612409
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)

AUTHORS
TITLE
JOURNAL
COMMENT

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 1105.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DC010AE12QP1&cluster=1105.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DC010AE12QP1.

FEATURES
source
1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC010Y123"
/tissue type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 305 a 241 c 340 g 275 t 40 others
ORIGIN

Alignment Scores:
Pred. No.: 1 51e-136 Length: 1201
Score: 1501.00 Matches: 281
Percent Similarity: 99.30% Conservative: 3
Best Local Similarity: 98.25% Mismatches: 2
Query Match: 97.09% Indels: 1
DB: 13 Gaps: 0

US-10-054-313-1 (1-286) x BX394121 (1-1201)

QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 99 YTGAGTGGCTTCTGTTCTGGCCACAGATGCGCTTGGCCGCTTGGCCGCGCGC 158
QY 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgArgGlyValPhe 40
Db 159 GGTCTTCGGGGTTCGGAGTGTCTATGCCGTGAGGAGGGCCGCAAGACGGGCTTT 218
QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60
Db 219 CTGACCTGGAATGAGTGCAGACACAGGTGGACCGGTTTCTCTGCCAGATTAAAGAG 278
QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 279 TTGGCCACAGAGATGAGGCTGGGCTTTGTGAGGAATCTCCAGCCGCGAGATTCA 338
QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 339 GAAGGGCATGAAATCAATCAATGACACAGAAATCGAGCGCGAAAGT-AGCAAGCGACTCCGT 397
QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTrpAlaLysHisMetLysPro 120
Db 398 GAGCCACTGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 457
QY 121 SerValGluProAlaProValSerArgAspThrPheSerTrpMetGlyAspPheVal 140
Db 458 AGCGTGAGCGCGCGCTCCAGTTAGCAGACACAGCTTTTCTTACATGGGAGACTTCGTC 517
QY 141 ValValThrTrpAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
Db 518 GTCTCTACATGATGCTGTCTCCAGTAATGGGCGTGAAGCGCGCGAGCAGGAATC 577
QY 161 GlyValThrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180

Db 578 GCGTTTACTGGGGCCAGGCCATCTTTAAATGTAGGCATTAGACTTCTCTGGCGGCAG 637
QY 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
Db 638 ACAACCAAGAGAGCGGAATTCATGAGCTCTCAAGCCATTGACAAAGAACTCAA 697
QY 201 AsnIleAsnLysLeuValLeuTrpThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 698 AACATCAATAAATGCTTCTGTATACAGACAGTATGTTTACGATAAATGGTATACTAAC 757
QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
Db 758 TGGTTCAGAGTTGGAAGAAAATGGGTGAAGACAAAGTGCAGGAAAGAGGTGATCAAC 817
QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db 818 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCAGGGGATGGACATTCAGTGGATGCAT 877
QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db 878 GTTCTCTGTCATTCGGGATTTATAGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA 937
QY 281 AlalysGlnSerGluAsp 286
Db 938 GCTAAACAATCGGAAGAC 955

RESULT 2
AL560874 1201 bp mRNA linear EST 31-MAY-2003
LOCUS AL560874 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0DL005YJ14 5-PRIME, mRNA sequence.
ACCESSION AL560874 GI:31285003
VERSION AL560874.2
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 15, 2001 this sequence version replaced gi:12907756.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 1105.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DL005DE07QP1&cluster=1105.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DL005DE07QP1.

FEATURES
source
1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL005YJ14"
/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 300 a 248 c 358 g 259 t 36 others
ORIGIN

Alignment Scores:
Pred. No.: 2,966-136 Length: 1201
Score: 1498.00 Matches: 281
Percent Similarity: 98.95% Conservative: 2
Best Local Similarity: 98.25% Mismatches: 3
Query Match: 96.90% Indels: 1
DB: 9 Gaps: 0

US-10-054-313-1 (1-286) x AL560874 (1-1201)

Qy 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 142 ATGAGCTGGCTTCGTCTCTGCGCCACAGAGTCGGCTTGGCGCGCTTGGCTTGGCGCGC 201

Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgGlyThrGlyValPhe 40
Db 202 GCGCTCTCGGGTTCGGATGTTCTATGCCGTGAGAGGGCGCGCAAGCCGGGGTCTTT 261

Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db 262 CTGACCTGGATGAGTGCAGACACAGGTGGCCGGTTTCTGCTGCCAGATTAAAGAG 321

Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 322 TTGGCCACAGAGGATGAGCGCTTGGCTTGTGCGAAATCTGCAAGCCGGAAGTTCA 381

Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 382 GAAGGGCATGAAATCAACATGCAAGAAATCGAGGGCGAAGC-AGCAAGCGACTCCGT 440

Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 441 GAGCCACTGATGGAGATGGACATGAAAGCGCAGAGCGGTATGCAAGACACATGAAGCG 500

Qy 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 501 ACCGTGGAGCGCGCCCTCCAGTTAGCAGACACAGTTTTCCTACATGGGAGACTTCGT 560

Qy 141 ValValTyrThrAspGlyCysSerSerAsnGlyArgGlyProArgAlaGlyIle 160
Db 561 GTCGCTACACTGATGGCTGCTGCTCCAGTAATGGCGGTAGAGCGCGGAGCAATC 620

Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db 621 GCGGTTTACTGGGGGCGAGCCATCCTTTAAATGTAGGCATTAGACTTCTCGGGCGGAG 680

Qy 181 ThrAsnGlnArgAlaGluIleHisAlaLysLysAlaLysGluGlnAlaLysThrGln 200
Db 681 ACACCAAGAGCGGAAATTCATGCGAGCTGCAAGCCATTGAACAGCAAGCAACTCA 740

Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 741 AACATCAATAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC 800

Qy 221 TrpValGlnGlyTrpLysLysAsnGlyTyrLysThrSerAlaGlyLysGluValIleAsn 240
Db 801 TGGGTTCAAGTTGGAGAAAATGGGTGGAAGACAGTGCAGGGAAGAGGTGATCAAC 860

Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db 861 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCAGGGGATGGACATTCAGTGGATGCAT 920

Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db 921 GTTCTCGGTTCATTGGGGATTTATAGGCAATGAAGAAGCTTGACAKATTAGCCAGAGAAGA 980

Qy 281 AlaLysGlnSerGluAsp 286
Db 981 GCTAAACATCGGAGAC 998

RESULT 3
BM541420 1054 bp mRNA linear EST 20-FEB-2002
LOCUS

DEFINITION AGENCOURT_6493820 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5521285
5', mRNA sequence.
ACCESSION BM541420
VERSION BM541420.1 GI:18770049
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1054)
TITLE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-f@mail.nih.gov
Tissue Procurement: ARCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM2186 row: k column: 14
High quality sequence stop: 713.
Location/Qualifiers
source
1..1054
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5521285"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 71"
/notes="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb." 214 t 4 others
BASE COUNT 274 a 239 c 323 g 214 t
ORIGIN
Alignment Scores:
Pred. No.: 4,036-132 Length: 1054
Score: 1455.00 Matches: 278
Percent Similarity: 97.24% Conservative: 4
Best Local Similarity: 95.86% Mismatches: 4
Query Match: 94.11% Indels: 4
DB: 12 Gaps: 0
US-10-054-313-1 (1-286) x BM541420 (1-1054)
Qy 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 45 ATGAGCTGGCTTCTGTTCCTGGCCACAGAGTCGGCTTGGCGCGCTTGGCTTGGCGCGC 104
Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgGlyThrGlyValPhe 40
Db 105 GCGCTCTCGGGTTCGGATGTTCTATGCCGTGAGAGGGCGCGCAAGCCGGGGTCTTT 164
Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db 165 CTGACCTGGATGAGTGCAGACACAGGTGGCGGTTCCTCTGCCAGATTAAAGAG 224
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 225 TTGGCCACAGAGGATGAGCGCTTGGCTTGTGCGAAATCTGCAAGCCGGAAGTTCA 284
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 285 GAAGGGCATGAAATCAACATGCAAGAAATCGAGGGCGAAGCCAGCAAGCGACTCCGT 344
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 345 GAGCACTGGATGGATGGACATGAAGCGCAGAGCGCTTATGCAAGACCATGAAGCCG 404

121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
 Db 405 AGCGTGGAGCGCGCGCTCCAGTTAGCAGACAGCGTTTCTTACATGGGAGCTTCTC 464
 Qy 141 ValValTyrThrAspGlyCysCysSerSerArgGlyValProArgAlaGlyIle 160
 Db 465 GTCTGTTACACTGATGGCTGCTCTCCAGTAATGGCGTAGNAGGCCCGGAGGATC 524
 Qy 161 GlyValTyrTyrProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
 Db 525 GCGGTTTACTGGGCGCGCGCATCTTTAAATGTAGGCATTAGACTTCTCTGGCGGAG 584
 Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
 Db 585 ACAACCAAGAGCGGAATTCATGAGCCTCCAAAGCCATTGAACCAAGCAAGACTCA 644
 Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
 Db 645 AACATCAATAAAGTGTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC 704
 Qy 221 TrpValGlnGlyTyrLysLysAsnGlyTyrLysThrSerAla-GlyLysGluValIleA 240
 Db 705 TGGGTTCAAGGTGGAGGAAATAGGTTGGAGGAGGAGGAGGAGGAGGAGGAGGAG 764
 Qy 240 snlyGluAspPheValAlaLeuGluArgLeuThrGln-GlyMetAspIleGlnTrpMet 259
 Db 765 ACAAGAGGACTTGTGGGCTGGAGAGGCTTACCCAGGCGGATGACATTCAGTGGATG 824
 Qy 260 HisValPro-GlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgG 279
 Db 825 CATGTTCCCTGGGTCATTCGGGATTTATAGGCCATGAAGAACTGACAGATTAGCCAGGAA 884
 Qy 279 uGlyAlaLysGlnSerGluAsp 286
 Db 885 AGAGGCTAAACCATCGGAGAC 906

RESULT 4
 BM810651
 LOCUS
 DEFINITION AGENCOURT_6580549 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5453776
 5', mRNA sequence.
 ACCESSION BM810651
 VERSION BM810651.1 GI:19127474
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1060)
 NIH-MGC <http://mgi.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 CONTACT Contact: Robert Strausberg, Ph.D.
 Email: csapbs@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: LICM1947 row: n column: 17
 High quality sequence stop: 725.
 Location/Qualifiers
 1..1060
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /cclone="IMAGE:5453776"
 /tissue_type="astrocytoma grade IV, cell line"
 /lab_host="D310B (phage-resistant)"
 /clone_lib="NIH_MGC_98"
 /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:

EcORI; cDNA made by oligo-dT priming. Directionally
 cloned into EcORI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH MGC library."

BASE COUNT 265 a 246 c 332 g 213 t

ORIGIN

Alignment Scores:

Pred. No.: 3,09e-131 Length: 1060
 Score: 1446.00 Matches: 274
 Percent Similarity: 96.86% Conservative: 4
 Best Local Similarity: 95.47% Mismatches: 8
 Query Match: 93.53% Indels: 2
 DB: 12 Gaps: 0

US-10-054-313-1 (1-286) x BM810651 (1-1060)

Qy 1 MetSerTyrLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArg 20
 Db 58 ATGAGCTGGCTTCTGTCTGGCCACAGAGTCGCTTGGCCGCTTGGCCCTCCGCCGC 117
 Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgGlyThrGlyValPhe 40
 Db 118 GCGTCTCGCGGTTCCGGGATGTTCTATGCCGTGAGGAGGGGGCCGCAAGACCGGGTCTTT 177
 Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60
 Db 178 CTGACCTGGAAATGAGTGCAGAGCACAGGTGGACGGTTTCTCTGCCAGATTAGAAG 237
 Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerLaserProGluValSer 80
 Db 238 TTTGCCACAGAGGATGAGGCTGGGCTTGTCTCAGGAAATCTGCAGCCCGGAGTTTCA 297
 Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
 Db 298 GAAGGGCATGAATAATCAACATGGCAAGAATCGGAGCGGAAGCCAGCAAGGACTCCGT 357
 Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
 Db 358 GAGCCACTGGATGGAGATGGACATGAAGCGCAGAGCCGTATGCAAGACACATGAAGCCG 417
 Qy 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
 Db 418 AGCGTGGAGCGCGCGCTCCAGTTAGCAGAGACACGTTTCTTACATGGGAGACTTCTC 477
 Qy 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgGlyLysProArgAlaGlyIle 160
 Db 478 GTCGCTACACTGATGGCTGCTCTCCAGTAATGGCGTAGAGGCCCGGAGGAGGATC 537
 Qy 161 GlyValTyrTyrProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
 Db 538 GCGGTTTACTGGGCGCGCGGCATCTCTTTAAATGTAGGCATTAGACTTCTCTGGCGGAG 597
 Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
 Db 598 ACAACCAAGAGCGGAATTCATGAGCCTCCAAAGCCATTGAACCAAGCAAGACTCAA 657
 Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
 Db 658 AACATCAATAAAGTGTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC 717
 Qy 221 TrpValGlnGlyTyrLysLysAsnGlyTyrLysThrSerAlaGlyLysGluValIleAsn 240
 Db 718 TGGGTTCAAGGTGGAGAA-ATGGGTGGAGAGCAAGTGCAGGGAAGAGGTGATNCAC 776
 Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
 Db 777 AAAGAGGACTTGTGGCACTGGAGAGGCTTACCAGGCGGATGGACATTCAGTGGATGCAT 836
 Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArg-LeuAlaArgGlu 280


```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4905246"
/tissue_type="zabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 17"
/notes="Organ: muscle; Vector: pOTB7; Site 1: EcoRI;
Site 2: XhoI; cDNA made by clico-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGACAGG(G). Size-selected >500bp
for average insert size 1.8Kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 255 a 204 c 295 g 190 t

ORIGIN

Alignment Scores:

| | | | |
|------------------------|-----------|---------------|-----|
| Pred. No.: | 2,566-128 | Length: | 944 |
| Score: | 1415.50 | Matches: | 271 |
| Percent Similarity: | 96.18% | Conservative: | 6 |
| Best Local Similarity: | 94.10% | Mismatches: | 8 |
| Query Match: | 91.56% | Indels: | 4 |
| DB: | 12 | Gaps: | 1 |

US-10-054-313-1 (1-286) x BG829086 (1-944)

QY 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20

Db 4 ATGAGTGGCTTCTGTCTTCTGCGCCACAGAGTGGCTTGGCGGCTTGGCGGCGGCGG 63

QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgGlyThrGlyValPhe 40

Db 64 GGCTCTCGGGTTCGGGATGTCTATATGCGGTGAGGAGGGCGGCGGAGACCGGGGTCTTT 123

QY 41 LeuThrTrpAsnGlnCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60

Db 124 CTGACCTGGATGATGTCAGAGACAGTGGACCGGTTTCTTGTCTGCCAGATTTAAGAG 183

QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80

Db 184 TTTGCCACAGAGGATGAGGCTGGGCTTTGTTCAGGAATCTGCAAGCCGCGGAAGTTTCA 243

QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaValProGlyLysArgLeuArg 100

Db 244 GAAGGGCATGAAATCAACATGACCAAGAAATCGAGCGGAGAACCCAGCAGGACTCCGT 303

QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120

Db 304 GAGCCACTGGATGAGATGACATGAAAGTGCAGAGCGGTATGCCAAAGCACATGAAGCG 363

QY 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140

Db 364 AGCGTGGAGCGCGCGCTCCAGTTAGCAGAGACACGTTTTTCTTACATGGAGACTTCGTC 423

QY 141 ValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160

Db 424 GTGCTGTACACTGATGCTGCTCTCCAGTAATGGCGTAGAAGCGCGGAGCAGGAATC 483

QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180

Db 484 GCGCTTTACTGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 543

QY 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCys-LysAlaIleGluGlnAlaLysThrGln 200

Db 544 ACAACCCAAAGAGCGGGAATTCATGCAGCTTCGCAAGGCCATTTGAACAAGCAAGACTCA 603

QY 200 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAs 220

Db 604 AAACATCAATAAATGCTTCTGTATACAGACAGTATGTTTACGATAAATGGTAACTAA 663

QY 220 nTrpValGlnGlyTrpLysLysAsnGlyTyrLysThrSerAlaGlyLysGluValIleAs 240

664 CTGGTTCAAGTTTGAAGAAAATGGTGGAGACAAAGTCAGGCAAGAGGTGATCAA 723

QY 240 nLysGluAspPheValAla-LeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetH 260

Db 724 CAAAGAGGACTTTTGTGGCAGCTTGGAGAGGCTTACCAGGGGATGGACATTCAGTGGATGC 783

QY 260 isValProGlyHisSerGlyPheIleGlyAsnGluAlaAlaAspArgLeuAlaArgGluG 280

Db 784 ATGTACCGGTCATCGGGATTAT--AGGCATGAAGAGTGCACAGATTAGCCAGAGAAG 840

QY 280 lylalysGlnSerGluAsp 286

Db 841 GAGTTAA-CAATCGAAGAC 859

RESULT 7

EX416456 1196 bp mRNA linear EST 15-MAY-2003

LOCUS EX416456 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone

DEFINITION CSODA007Y012 5-PRIME, mRNA sequence.

ACCESSION EX416456

VERSION EX416456.1 GI:30763629

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1196)

AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 1105.r For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CSODA007BH06QPI&cluster=1105.r. Contact :

Feng Liang Email : fliang@lifetech.com URL : Invitrogen Corporation 1600

http://fulllength.invitrogen.com/

Faraday Avenue Genoscope sequence ID : CSODA007BH06QPI.

Location/Qualifiers

1. 1196

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CSODA007Y012"

/tissue_type="NEUROBLASTOMA"

/clone_lib="Homo sapiens NEUROBLASTOMA"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo(dT) primer. Five prime end enriched,

double-strand cDNA was digested with Not I and cloned into

the Not I and EcoRV sites of the pCMVSPORT 6 vector.

Library was not normalized."

BASE COUNT 300 a 247 c 345 g 269 t 35 others

ORIGIN

Alignment Scores:

| | | | |
|------------------------|-----------|---------------|------|
| Pred. No.: | 3,02e-127 | Length: | 1196 |
| Score: | 1406.00 | Matches: | 277 |
| Percent Similarity: | 97.55% | Conservative: | 2 |
| Best Local Similarity: | 96.85% | Mismatches: | 7 |
| Query Match: | 90.94% | Indels: | 4 |
| DB: | 13 | Gaps: | 0 |

US-10-054-313-1 (1-286) x BX416456 (1-1196)

QY 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20

Db 125 ATGAGCTGGTCTGTCTTCTTCTGCGCCACAGAGTCGCTTGGCGGCGGCGGCGGCGG 184

Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgGlyThrGlyValPhe 40
Db 185 GGCTCTCGCGGTTTCGGATGTTCTATGCCGTGAGGAGGCGCGCAAGACCGGGGCTTT 244
Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db 245 CTGACCTGGATGATGTCAGAGCAGACAGTGGACCGGTTTCTCTGCTGCACATTTAAGAAG 304
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 305 TTTGCCACAGAGGATGAGCGCTGGCGCTTTGTACAGAAATCTGCAAGCCGGAAGTTTCA 364
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 365 GAAGGGCATGAAATACACATGACAGATCGAGGCGGAAGC-ACMAGCGACTCCGT 423
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 424 GAGCCACTGGATGAGATGGACATGAAAGCCAGACGCGGTATGCAAGACCATGAAGCCG 483
Qy 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 484 AGCGTGGAGCGCGCGCTCCAGTTAGCAGACACAGTTTCTTACATGGGAGACTTCGTC 543
Qy 141 ValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160
Db 544 GTCGCTACACTGATGGCTGCTGCTCCAGTAAATGGGCTAGAGCGCGCAGCAGGAATC 603
Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLeuArgLeuProGlyArgGln 180
Db 604 GCGCTTACTGGGGCCAGGCCATCCCTTAAATGTAGGCATATAGACTTCTCGGCGCGAG 663
Qy 181 ThrAsnGlnArgAlaGluLeuHisAlaAlaCysLysAlaLeuGluGlnAlaLysThrGln 200
Db 664 AAAAAACCAAGCGGAATTCATGCGCTTGCAGCCATGCAAGCCATTGAACGAAGCAACTCAA 723
Qy 201 AsnLeuAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 724 AACATCAATAAAGTGGTCTGTATACAGACAGTATGTTTACGATAAATGGTATTAACTAAC 783
Qy 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
Db 784 TGGGTTCAGGTTTGAAGAA- AATGGTGGGAAGCAAGTGCAGGGAAGAGGTGATCAC- 841
Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db 842 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCAGGGGATGGACATTCAGTGGATGCAT 901
Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGluGly 280
Db 902 GTTCT-GGTCTATCGGGATTTATAGGCAATGAAGAAGCTGACATAGCCAGAGAAGGA 960
Qy 281 AlaLysGlnSerGluAsp 286
Db 961 GCTAAACATCGGAAGAC 978

RESULT 8
BUI75613
LOCUS
DEFINITION BUI75613 893 bp mRNA linear EST 04-SEP-2002
AGENCY 7935024 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6009671
5', mRNA sequence.
ACCESSION BUI75613
VERSION BUI75613.1 GI:22689597
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eukarya; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 893)
NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov
Tissue Procurement: DCTD/DP/Carzdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LAM13196 row: h column: 24
High quality sequence stop: 672.
Location/Qualifiers
1. .893
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6009671"
/issue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_68"
/note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 237 a 197 c 279 g 178 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 1,91e-121 Length: 893
Score: 1345.00 Matches: 270
Percent Similarity: 95.14% Conservative: 4
Best Local Similarity: 93.75% Mismatches: 4
Query Match: 87.00% Indels: 10
DB: 13 Gaps: 0
US-10-054-313-1 (1-286) x BUI75613 (1-893)

Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 27 ATGAGCTGGCTTCTGTCTCTGCCACACAGTGGCTTGGCGGCTTGCCTGCCCGCCG 86
Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
Db 87 GCGTCTCGCGGTTTCGGGATGTTCTATGCGGTGAGAGGGCGCGCAAGACCGGGCTTT 146
Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db 147 CTGACCTGGAATGATGTCAGAGCAGACAGTGGACCGGTTTCTCTGCCAGATTTAAGAAG 206
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 207 TTTGCCACAGAGGATGAGCGCTTGGGCTTTTCAGGAAATCTTCAAGCCCGCGAAGTTTCA 266
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 267 GAAGGGCATGAAATCAACATGACAAAGAAATCGAGGCGAAAGCCAGCAAGCGACTCCGT 326
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 327 GAGCCACTGGATGAGATGACATGAAGCGCAGACCGCTATGCAAGACCATGAAGCCG 386
Qy 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 387 AGCGTGGAGCGCGCTCCAGTTACAGACACACGCTTTTCTTACATGGGAGACTTCGTC 446
Qy 141 ValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160
Db 447 GTCGCTACACATGATGGCTGCTCCAGTAATGGCGGTAGAGCGCGCGAGCAGGAATC 506
Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLeuArgLeuProGlyArgGln 180
Db 507 GCGGTTTACTGGGGCGCGGCCATCCCTTAAATGTAGGCATTAGACTTCTTGGCGCGAG 566
Qy 181 ThrAsnGlnArgAlaGluLeuHisAlaAlaCysLysAlaLeuGluGlnAlaLysThrGln 200


```

|||||
567 ACAACCAAGAGAGGAAATTCATGACGCTGCAAGCAATTAAGCAAGCAAGCACTCAA 626
QY |||||||
201 AsnIleAsnLysLeuValIleuTyThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db |||||||
627 AACATCAATAAAGTCTGTATACACACAGTAGTATGTTACGATAAATGGTAACTAAC 686
QY |||||||
221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
Db |||||||
687 TGGGTTCAAGGTTGGAGAAATGGTGGAGCAAGTGCAGGGAAGAGGTGATCAAC 746
QY |||||||
241 LysGluAspPhe-ValAlaLeu-GluArgLeuThrGlnGly--MetAspIleGlnTrp-M 259
Db |||||||
747 AAAGAGGACTTTGGTGGCACTGGGAGAACTTACCCAGGGGATGGGACATTCAGTGGGA 806
QY |||||||
259 eHis-ValProGly-His-SerGlyPheIle-GlyAsn-GluGluAlaAspArgLeuAl 277
Db |||||||
807 TGCATGGTTCCTGGNCCATTCGGGATTTATAGGCAATGGAAGAACTGACAGATTAGC 866
QY |||||||
277 aArg 278
Db |||
867 CCAG 870

RESULT 9
BG831321
LOCUS
DEFINITION
602766160F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4908326 5',
mRNA sequence.
ACCESSION
BG831321
VERSION
BG831321.1 GI:14178908
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 949)
AUTHORS
NIH-MGC http://img.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ARCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM184 row: 9 column: 15
High quality sequence stop: 807.
Location/Qualifiers
1..949
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4908326"
/tissue_type="epithelioid carcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_42"
/notes="Organ: pancreas; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Saratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 255 a 195 c 303 g 196 t
ORIGIN

```

Alignment Scores:

Pred. No.: 6.4e-121 Length: 949

```

Score: 1340.00 Matches: 257
Percent Similarity: 93.93% Conservative: 6
Best Local Similarity: 91.79% Mismatches: 15
Query Match: 86.68% Indels: 3
DB: 12 Gaps: 0

US-10-054-313-1 (1-286) x BG831321 (1-949)

QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 30 ATGAGCTGGCTTCTGTTCTGGCCACACAGATCGCCTTGGCCGCTTGGCCCTGCCCGC 89
QY 21 GlySerArgGlyPheGlyMetPheTyAlaValArgArgGlyArgGlyThrGlyValPhe 40
Db 90 GGCCTCTCGCGGGTTCGGGATGTTCTATCGCTGAGGAGGGCGCGAAGCCGGGCTTT 149
QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db 150 CTGACCTGGAATCAGTGCAGAGCACAGTGGAGCCGCTTCTGCTGCCAGATTTAAGAAG 209
QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 210 TTTGCCACAGAGATGAGCCCTGGGCTTTGTTCAGGAAATCTGCAAGCCCGGAAGTTTCA 269
QY 81 GluGlyHisGluAsnGlnHisGlyGlnGlySerGluAlaLysProGlyLysArgLeuArg 100
Db 270 GAAGGGCATCAAAATCAACATGACCAAGATCGAGGCGGAAGCCAGCAAGCACTCCGT 329
QY 101 GluProLeuAspGlyAspGlyHisGlySerAlaGlnProTyThrAlaLysHisMetLysPro 120
Db 330 GAGCCACTGGATGGAGATGGACATGAAGCGCAGAGCCGCTATCAAGACACATGAAGCCG 389
QY 121 SerValGluProAlaProProValSerArgAspThrPheSerTyMetGlyAspPheVal 140
Db 390 AGCGTGAGCGCGGCGCTCCAGTTAGCAGACACACGTTTCTTCATACATGGAGACTTCGTC 449
QY 141 ValValTyThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
Db 450 GTCGCTACACTGATGCTGCTGCTCAGTAATGGCGGTAGAAAGCGCGCGAGCAAGCAATC 509
QY 161 GlyValTyThrProGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db 510 GCGCTTTACTGGGCGCGCGGCGCTCCCTTTAAATGTAGGCATTAGACTTCTCTGGCGGCGAG 569
QY 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
Db 570 ACAACCAAGAGCGGAAATTCATGACGCTGCAAGCCATTCGAAAGCAAGCAAGCAAGCACTCAA 629
QY 201 AsnIleAsnLysLeuValIleuTyThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 630 AACATCAATAAAGTCTGTATACACACAGTAGTATGTTACGATAAATGGTAACTAAC 689
QY 221 TrpValGlnGlyTrp-LysLysAsn-GlyTrpLysThrSerAlaGlyLysGluValIleA 240
Db 690 TGGGTTCAAGGTTGGGAGAAATGGTGGAGACAGCAAGTGCAGGCGGCAAGAGGTGATC 749
QY 240 snLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMeth 260
Db 750 AAAAAGAGGACTTTGTGGGACTGGAGAGGTT-ACCAGGGATGGGACATTCAGTGGATGC 808
QY 260 isValProGlyHisSerGlyPheIleGlyAsnGluAlaAspArgLeuAlaArg 278
Db 809 ATGCTCGGACATCCCGGATTAATAGGAATGAGGAAGTGAAGGATTAGCGAGA 864

RESULT 10
CA775004
LOCUS
DEFINITION
765 bp mRNA linear EST 03-DEC-2002
ip11906.y1 HR85 islet Homo sapiens cDNA clone IMAGE:6216755 5',
similar to TR:060523 060523 RIBONUCLEASE H TYPE II. [2] TR:060857
// mRNA sequence.
CA775004
CA775004.1 GI:26012477
ACCESSION
CA775004
KEYWORDS
EST.

```

```

SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 765)
AUTHORS    Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
            Lemishka, L., Scarce, M., Bresnelli, J., Gradwohl, G., Clifton, S.,
            Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,
            Schmitt, A., Theising, B., Ritter, B., Ronko, I., Bennett, J., Cardenas
            M., Gibbons, M., McCann, R., Cole, R., Teagareishvili, R., Williams, T.,
            Jackson, Y., and Bowers, Y.
TITLE      Endocrine Pancreas Consortium
JOURNAL     Unpublished
COMMENT     Other ESTs: ip1lg06.xl
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 469.
FEATURES    Location/Qualifiers
             1..765
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="IMAGE:6216755"
             /issue_type="Purified pancreatic islet"
             /lab_host="DH10B"
             /clone_lib="HR85 islet"
             /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
             NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
             Size-selected on agarose gel. Average insert size ~1kb. 5'
             XhoI site was destroyed after directional cloning.
             Amplified once. Contact information: Hiroshi Inoue, MD,
             Metabolism Div. (Alan Permutt Lab), Washington University
             School of Medicine, Box 8127, 660 South Euclid Ave., St.
             Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
             314-362-1916, Fax: 314-747-2692."
BASE COUNT 202 a 177 c 237 g 149 t
ORIGIN
Alignment Scores:
Pred. No.:      5,91e-116      Length:      765
Score:          1288.00      Matches:      237
Percent Similarity: 99.17%      Conservative: 3
Best Local Similarity: 97.93%      Mismatches: 2
Query Match:      83.31%      Indels:      0
DB:              14          Gaps:      0

US-10-054-313-1 (1-286) x CA775004 (1-765)
QY      1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db      40 ATGAGCTGGCTTCGTCTCTGCGCCACAGAGTCGCGCTTGCGCGCTTGCGCGCGC 99
QY      21 GlySerArgGlyPheGlyMetPheTrpAlaValArgGlyArgGlyThrGlyValPhe 40
Db      100 GCGCTCGCGGGTTCGGATGTTCTATCCGTCGAGAGCGCGCGCAACCGGGGTCTTT 159
QY      41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPhePzoAlaAlaArgPheLysLys 60
Db      160 CTGACTGGAGTGAAGTCAGACACAGAGTCGGACCGGTTCTCTGTCGCCAGATTAAAGAG 219
QY      61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80

```

220 TTTGCCACAGAGATGAGGCCTGGCCCTTTGTTCAGAAATCTGCAAGCCCGGAAGTTTCA 279

81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100

280 GAAGGCGATGAAATCAACATCGACAGAATCGGAGGCGAAGCCAGCAGGACTCCGT 339

101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTrpAlaLysHisMetLysPro 120

340 GAGCCACTGGATGGAGATGGACATGAAGCGCAGAGCCGTATGCAAGACACATGAAGCG 399

121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140

400 AGCGTGGAGCGCGCGCTCCAGTTAGCAGAGACACGTTTTCTACATGGGAGACTTCGTC 459

141 ValValTrpThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160

460 GTGCTCTACACTGATGGCTGCTGCTCCAGTAATGGCGGTAGAGCGCGCAGCAGGAATC 519

161 GlyValTrpTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180

520 GCGTTTACTGGGCGCGAGCCATCCTTTAATGTAGGCATTAGACTTCTCTGGCGGCGAG 579

181 ThrAsnGlnArgAlaGluIleHisAlaLysLysAlaLysLysAlaLysLysThrGln 200

580 ACAACCAAGAGCGGAAATTCATGAGCGCTGCAAGCCATTGAAACAAGCAAGACTCAA 639

201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220

640 ACATCAATAACTGTTCTGTATACAGACAGATGTTTACGATTAATGTTAATTAATAAC 699

221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240

700 TGGTTCAAGGTGGAGAAATAATGGTGGAAAGCAAGTCGAGGGAAGAGGTGATCCAC 759

241 LysGlu 242

760 AAAGAG 765

RESULT 11

BQ921568

LOCUS

DEFINITION BQ921568.1 935 bp mRNA linear EST 20-AUG-2002

AGENCOUNT 8922346 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6470333

5', mRNA sequence.

BQ921568

VERSION BQ921568.1 GI:223336599

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 935)

AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>

Plate: LLAM14000 row: 0 column: 06

High quality sequence stop: 563.

Location/Qualifiers

1..935

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6470333"

/tissue_type="leiomyosarcoma"

/lab_host="DH10B (phage-resistant)"

FEATURES

source

```

/clone_lib="NTH_MGC_71"
/note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."
BASE COUNT 235 a 211 c 301 g 182 t 6 others
ORIGIN

Alignment Scores:
Pred. No.: 3,77e-115 Length: 935
Score: 1281.00 Matches: 255
Percent Similarity: 91.1% Conservative: 4
Best Local Similarity: 91.73% Mismatches: 11
Query Match: 82.86% Indels: 8
DB: 13 Gaps: 0

US-10-054-313-1 (1-286) x BQ921568 (1-935)
QY 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
DB 98 ATGAGCTGGCTTCTGCTTCTGCCCCACAGAGTCGCTTGGCGGCTTGCCCTGGCGCGC 157
QY 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgGlyArgGlyThrGlyValPhe 40
DB 158 GGCCTTCGCGGTTTGGGATGTTCTATGCCGTGAGAGGGGCCGCAAGCCGGGTCTTT 217
QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
DB 218 CTGACCTGGAATGAGTGCAGACACAGGTGGACCGGTTTCTGCTGCCAGATTTAAGAAG 277
QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
DB 278 TTGGCCACAGAGATGAGGCTTGGGCTTTGTCAGGAATCTGCAAGGCCGGAATTCA 337
QY 81 GluGlyHisGluAsnGlnHisGlyGlnGlnSerGluAlaLysProGlyLysArgLeuArg 100
DB 338 GAAGGCGATGAAATCAACATGCACAAAGATCGGAGGCGAAAGCCAGCAGCTCGT 397
QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTrpAlaLysHisMetLysPro 120
DB 398 GAGCCACTGGATGGAGATGGATGAAGCGCAGAGCCGATGCAAGACACATGAAGCCG 457
QY 121 SerValGluProAlaProValSerArgAspThrPheSerTrpMetGlyAspPheVal 140
DB 458 AGCGTGGAGCGCGCTCCAGTTAGCAGACACAGTTTCTACATGGAGACTTCGTC 517
QY 141 ValValThrTrpAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaCylle 160
DB 518 GTGCTTACACTGATGGCTGCTCTCCAGTAATGGCGTAGAAGCGCGGAGCAAGATC 577
QY 161 GlyValThrTrpGlyProGlyHisProLeuAsnValGlyLeuArgLeuProGlyArgGln 180
DB 578 GGGGTTTACTGGGGCGCGGCCATCCTTTANATGTAGGCATTAGACTTCTGGCGGCGAG 637
QY 181 ThrAsnGlnArgAlaGluLeuHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
DB 638 ACAACCAAGACGCGAAATTCATGAGGCTCCAAAGCCATTGAAACAAGAACTCAA 697
QY 201 AsnIleAsn-LysLeuValLeuTrpThrAspSerMetPheThrIleAsnGlyLysThrAs 220
DB 698 AACATCAATAAACTGGTCTGTATACACAGATGTTTACGATAAATGGTATAACTAA 757
QY 220 nTrpValGlnGlyTrpLys-LysAsnGlyTrpLysThrSerAla--GlyLysGluValI 239
DB 758 CTGGNGTTCAGGTGGGAAGAAATATGGTGGGAAGACAGTGCANGGGGAAAGAGGTGAT 817
QY 239 eAsnLysGluAspPheValAlaLeu-GluArgLeuThrGlnGlyMet-AspIleGlnTrp 258
DB 818 CAACAAGAGAGACTTTGTGGCATCTGGAGAGCTTAACCCAGGGAATGGGACATCAGNGN 877
QY 259 MetHis-ValProGlyHis-SerGlyPheIleGlyAsn 270
DB 878 ATGATGGTTTCTTGGGCAATTCGGGATTAATAGGCCAT 915

```

```

RESULT 12
BE746490
LOCUS BE746490 779 bp mRNA linear EST 15-SEP-2000
DEFINITION 601579445F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3928402 5',
mrna sequence.
ACCESSION BE746490
VERSION BE746490.1 GI:10160482
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 779)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: DCTB/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LICM759 row: i column: 11
High quality sequence stop: 746.
Location/Qualifiers
1..779
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3928402"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_9"
/note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACAGAG(G). Size-selected for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 218 a 161 c 242 g 158 t
ORIGIN

Alignment Scores:
Pred. No.: 3,41e-111 Length: 779
Score: 1239.50 Matches: 238
Percent Similarity: 95.63% Conservative: 3
Best Local Similarity: 94.44% Mismatches: 8
Query Match: 80.17% Indels: 4
DB: 10 Gaps: 1

US-10-054-313-1 (1-286) x BE746490 (1-779)
QY 18 CysArgArgGlySerArgGlyPheGlyMetPheTrpAlaValArgGlyArgLysThr 37
DB 2 TGC CGCGCGG--TCTCGCGGTTTCGGGATGTTCTATGCCGTGAGGAGGGCGCCAGAAC 58
QY 38 GlyValPheLeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArg 57
DB 59 GGGGTCTTTCGACCTGGAATGATGTCAGAGCACAGTGGACCGGTTTCTGCTGCCAGA 118
QY 58 PheLysLysPheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerPro 77
DB 119 TTTAAGAAGTTTGGCCACAGAGGATGAGCGCTGGGCGCTTGTCAAGAAATCTGCAAGCCG 178
QY 78 GluValSerGluGlyHisGluAsnGlnHisGlyGlnGlnSerGluAlaLysProGlyLys 97
DB 179 GAAGTTTCAGAAAGGCGCATGAAATCAACATGGACAGAAATCGGAGGCGAAAGCCAGCAAG 238
QY 98 ArgLeuArgGluProLeuAspGlyAspGlyHisGluSerAlaGlnProTrpAlaLysHis 117

```



```

Db      727 TCGGTTCCAGGCTTGAAGCAANAATGGGTGGNAGACNAGTGCAGGGGAAGAGGTGATC 786
Qy      240 Asn-LysGluAspPhe 244
      ::::|||||
Db      787 CACAAAGAGGACTTT 802

RESULT 14
LOCUS   BM451915
DEFINITION  BM451915 916 bp mRNA linear EST 05-FEB-2002
VERSION  AGENCOURT_6397386 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5492781
KEYWORDS  S', mRNA sequence.
SOURCE    BM451915.1 GI:18500955
          EST.
          Homo sapiens (human)
          ORGANISM
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 916)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgraphs@mail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12115 row: g column: 22
            High quality sequence stop: 626.

FEATURES
    source
        1..916
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:5492781"
            /tissue_type="retinoblastoma"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH MGC 67"
            /notes="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
            Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
            Average insert size 1.75 kb. Library constructed by Life
            Technologies."
BASE COUNT  223 a 229 c 282 g 180 t
ORIGIN
Alignment Scores:
Pred. No.: 2..2e-109 Length: 916
Score: 1222.00 Matches: 247
Percent Similarity: 89.08% Conservative: 6
Best Local Similarity: 86.97% Mismatches: 16
Query Match: 79.04% Indels: 15
DB: 12 Gaps: 4

US-10-054-313-1 (1-286) x BM451915 (1-916)
Qy      1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db      75 ATGAGCTGGCTTCTGTTCCTGGCCCAAGAGTGCCTTGGCCCTTGGCCCTGGCCGCGC 134
Qy      21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
Db      135 GGCTCTCGGGTTCGGATGTTCTATCGCGTGGAGGGGGCGGCGGACCGGGGTCTTT 194
Qy      41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db      195 CTGACCTCGGAATGAGTSCAGACACAGGTGGACCGGTTTCTGCTCCAGATTTAAGAG 254
Qy      61 PheAlaThrGluAspGluAlaTTPAlaPheValArgLysSerAlaSerProGluValSer 80

```

```

Db      255 TTTGCCACAGAGGATGAGGCTGGGCTTTTTCAGGAAATCTCAAGACCCGGAGTTTCA 314
Qy      81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
      ::::|||||
Db      315 GAAGGGCATGAAATCAACATGGACAGATCGAGCGGAAAGCCAGCAAGCATCCGT 374
Qy      101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
      ::::|||||
Db      375 GAGCCACTGGATGGAGATGGACATGAAGCGCAGACCGGTATCAAAAGCACATGAAGCCG 434
Qy      121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
      ::::|||||
Db      435 AGCGTGGAGCCGGCGCTCCAGTTAGCAGAGACACGTTTCTTACATGGAGACTTCGTC 494
Qy      141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
      ::::|||||
Db      495 GTGCTCTACACTGATGCTGCTGCTCTCCAGTAATGGCGTAGAAGCGCGGAGCAGGAATC 554
Qy      161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
      ::::|||||
Db      555 GCGGTTTACTGGGGGCGCAGGCCATCTCTTAAATGTAGGCATTAGACTTCTCTGGCGGCGAG 614
Qy      181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysGlyAlaIleGluGlnAlaLysThrGln 200
      ::::|||||
Db      615 ACAACCCANAGCGGGAATTCATGCAGCCTGCAAGCCATTGAACAAGCAAGACTCAN 674
Qy      201 AsnIleAsn-LysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAs 220
      ::::|||||
Db      675 AACATCCATAAACTGGTCTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAA 734
Qy      220 nTTP-ValGlnGly-TpLysLysAsn-GlyTTP-LysThrSerAla-GlyLysGluValI 239
      ::::|||||
Db      735 CTGGGTTTCCAGGTTTGGAGAAATAATGGGTGGAGACCAGTGCAGGGAAAGAGGTGA 794
Qy      239 leAsn---LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnT 258
      ::::|||||
Db      795 TTCACCAAGAAGAGGACTTTTGTGCGCACTGGAAGGCTCTACCCAGGG-----GGAT 845
Qy      258 rpMetHis-----ValProGlyHisSerGly---PheIleGlyAsnGluG 272
      ::::|||||
Db      846 GGACCTTTCCCGGGGATGCTTGTTCCTCCGGGCGCATCCGGGCATTTTATAGGCCAATGAAA 905
Qy      272 lu 272
      ::::
Db      906 AA 907

RESULT 15
LOCUS   BF984400
DEFINITION  602307923F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4399301 5',
          mRNA sequence.
ACCESSION  BF984400
VERSION    BF984400.1 GI:12387212
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
          ORGANISM
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 869)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgraphs@mail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10102 row: n column: 06
            High quality sequence stop: 691.

```

FEATURES
source

Location/Qualifiers

1. .869
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="IMAGE:4399301"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_88"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
oligo-dr primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 249 a 177 c 270 g 173 t
ORIGIN

Alignment Scores:

Pred. No.: 9.92e-109 Length: 869
Score: 1215.00 Matches: 231
Percent Similarity: 97.90% Conservative: 2
Best Local Similarity: 97.06% Mismatches: 4
Query Match: 78.59% Indels: 2
DB: 10 Gaps: 0

US-10-054-313-1 (1-286) x BF984400 (1-869)

| | | | |
|----|-----|--|-----|
| Qy | 47 | ArgAlaGlnValAspAtgPheProAlaAlaArgPheLysLysPheAlaThrGluAspGlu | 66 |
| Db | 3 | AGAGCACAGGTGGACCGGTTTCTGCTGCCAGATTATAGAAGTTTGGCCACAGAGGATGAG | 62 |
| Qy | 67 | AlaTrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGln | 86 |
| Db | 63 | GCCTGGGGCTTTGTCCAGAAATCTGCAAGCCGGAAGTTTCAGAAGGGCATGAAATCAA | 122 |
| Qy | 87 | HisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAsp | 106 |
| Db | 123 | CATGGACAAGAAATCGAGGCGAAGCCAGCAGGCGACTCCGTGAGCCACTGGATGGAGAT | 182 |
| Qy | 107 | GlyHisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaPro | 126 |
| Db | 183 | GGACATGAAGCGCAGACCGCTATGCAAGCACATGAAGCCGAGCGGTGGAGCGGCGCCT | 242 |
| Qy | 127 | ProValSerArgAspThrPheSerTyrMetGlyAspPheValValTyrThrAspGly | 146 |
| Db | 243 | CCAGTTAGCAGACACACGTTTCTCCATATGGGAGACTTCGTCGTCGTACACTGATGGC | 302 |
| Qy | 147 | CysCysSerSerAsnGlyArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyPro | 166 |
| Db | 303 | TGCTGCTCCAGTAATGGCGGTAGAGCGCGGAGCAGGAATCGGCGTTTACTGGGGGCG | 362 |
| Qy | 167 | GlyHisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGlu | 186 |
| Db | 363 | GGCCATCCTCTTTAAATGTAGGCATTAGACTTCTGGCGGCGCAGACAAACCAAGAGCGAA | 422 |
| Qy | 187 | IleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuVal | 206 |
| Db | 423 | ATTGATGAGCCCTGCAAGCCATTGACACAGCAAGACTCAAAACATCATTAATCTGTT | 482 |
| Qy | 207 | LeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLys | 226 |
| Db | 483 | CTGTATACAGACAGTATGTTACGATAAATGGTATAACTGCGGTTCAAGGTGGAAAG | 542 |
| Qy | 227 | LysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAla | 246 |
| Db | 543 | AAAAATGGGTGGAACACAAAGTCAGCGGAAAGAGGTGATCAACAAGAGAGACTTTGTGGA | 602 |
| Qy | 247 | LeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGly | 266 |
| Db | 603 | CTGGAGAGGCTTACCCA-GGGATGGACATTTCAGTGGATGATGTTCTCTGTTCAATCGGA | 661 |
| Qy | 267 | PheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly-AlaLysGln | 283 |

Db

662 TTTATAGCAATGAAGAGCTGACAGATTAGCCAGAGGAGGCTAAACAA 713

Search completed: December 17, 2003, 21:11:15
Job time : 2528 secs